Principled Explanations in Comparative Biomusicology – Toward a Comparative Cognitive Biology of the Human Capacities for Music and Language

Inaugural dissertation
to complete the doctorate from the Faculty of Arts and Humanities
of the University of Cologne
in the subject Musicology

presented by
Rie Asano

born on May 19th, 1986
in Tokyo, Japan

Cologne, Germany, April 10th, 2019

Primary supervisor
Prof. Dr. Uwe Seifert

Secondary supervisor
Prof. Dr. Cedric Boeckx

Tertiary supervisor
PD Dr. Doris Mücke

Table of contents

Acknowledgements

The seeds of this thesis were sown when I finished my magister thesis in 2011. Over almost eight years, this thesis experienced different periods and in each period there are a plenty of people who I would like to acknowledge here. Before beginning with the 'scientific' acknowledgements, first of all, I would like to say thanks to my families who believed in and supported me in each period of thinking and writing. Without their help, it was not possible for me to complete this thesis, so I am grateful to them, especially. Special thanks to Florian Römer, Kazuharu Asano, Keiko Asano, Eita Asano, Ursula Römer, Theo Römer, Gaby Ceh, and Rainer Ceh. I also would like to say thanks to my friends and colleagues who supported me, especially Alexandra Streck and Birgit Klemann.

The first period of my thesis was between 2011 and 2013 in which I was still very focused on comparative language-music research from theoretical linguistic and musicological perspectives. In this period, Noam Chomsky helped me a lot to understand his Minimalist Program. He always wrote back e-mails to answer my questions, which was very helpful. In addition, I would like to thank Jürgen Lenerz. Without his seminars, I would not have had opportunity to learn so much about generative linguistics, Minimalist Program, and biolinguistics. During the first period, I really enjoyed a discussion group with Samira Grabarz and Timo Varelmann on the modularity of mind. Without this discussion group, I might not have found the problems which I discussed in the current thesis. In addition, I would like to acknowledge two major critical comments. One was made by Ray Jackendoff on my at that moment still merge-centered view of language and music. The other was given by Isabelle Peretz to my heavily theoretical research strategy. Now, I am so grateful to those two critical comments which convinced me to extend my perspective. Especially, Ray Jackendoff keeps discussing now over eight years with me via e-mails and helped me continually with developing my idea.

The second period is a kind of transition period between 2013 and 2016 where I developed increasing interest in cognitive neuroscience of language and music. Cedric Boeckx was the main actor of this period who convinced me to do research more in light of biology. His paper *Biolinguistics: forays into human cognitive biology* influenced me a lot and provided me with new research problems. Inspired by *Comparative Biolinguistics Workshop* organized by Cedric Boeckx, I organized *Comparative Biomusicology Workshop* which is a milestone of the current thesis. I am very thankful to Cedric Boeckx, Tecumseh Fitch, Stefan Koelsch, Kai Vogeley, Ina Bornkessel-Schlesewsky, and Matthias Schlesewsky for having taken part in this

small, but very fruitful workshop, and discussed this new research framework with me. In this second period, too, I would like to acknowledge critical comments, which were this time provided by Michael A. Arbib. One of his critical comment dealt with the relationship between syntax and semantics, which I discussed in PART IV on the basis of the current thesis. In addition, I would like to say many thanks to Clemens Maidhof who helped me a lot in this second phase. Especially, I learnt much about empirical research and thinking. I also learnt how important it is to have a great research partner.

The third period is the phase between 2016 and 2019, where I developed the most of my ideas which I discussed in the current thesis. Especially, in this phase, I am very grateful of the support by Evolinguistics Project in Japan (MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas #4903 (Evolinguistics), Grant Number JP17H06379), especially Kazuo Okanoya and Koji Fujita, who provided me with many opportunities to present and discuss my work in an interdisciplinary environment. In addition, Genta Toya helped me a lot to understand computational modelling approach. I am also happy to have spent much time with discussing the notion of syntax with Volker Struckmeier. Special thanks also go to my students, especially Marvin Heimerich, who challenged my idea during the seminars with interesting questions. I really appreciated to discuss the materials with students because they see many new, unexpected things. I am also grateful of Sebastian Klaßmann and Timo Varelmann for comments on the penultimate draft of the current thesis.

My final set of thanks goes to my advisors. Doris Mücke was a lecturer in phonetics as I started my Magister degree in Germany. The lectures and seminars given by her and Martine Grice caused my interest in prosody and speech gesture which I hope to have linked a little bit to my idea of rhythmic syntax. I, once more, would like to thank Cedric Boeckx, this time, for being my advisor and helping me continuously in bad and good period. I really do not know how to express this gratitude in words. Last but not least, my huge thanks, of course, go to Uwe Seifert. I am especially grateful of him for giving me freedom to develop my own idea in my own way. I think, in the current scientific environment, it is very rare to have this freedom. In addition, Uwe caused my interest in computational neurocognitive modeling. I am especially thankful regarding this influence by Uwe as I will continue working on this topic in my future research work.

Introduction

1 In search of a comparative biological approach to cognitive systems

In his influential book *Biological Foundations of Language*, Eric Lenneberg (1967, p. 394) suggested the existence of "a biological matrix with specifiable characteristics" for each cognitive capacity. This book inspired a line of research programs investigating biological specifications of the language capacity. It made the term "biolinguistics", which was already introduced in 1950, resurface in an interdisciplinary meeting organized by Massimo Piattelli-Palmarini in 1974 (for historical reviews, see also Boeckx & Grohmann, 2007; Boeckx & Martins, 2016; Chomsky, 2005; P. T. Martins & Boeckx, 2016). Boeckx and Grohmann (2007, p. 2) regarded Lenneberg's research program as "biolinguistics in a strong sense" combing "linguistic insights and insights from related disciplines (evolutionary biology, genetics, neurology, psychology, etc.)", in contrast to "biolinguistics in a weak sense" referring to Chomsky's program to discover properties of grammar initiated in his book *Syntactic Structures* (Chomsky, 1957).

The term "biolinguistics" is often associated with the weak sense, especially with the *Minimalist Program* (Chomsky, 1995), and earned critics (e.g., Jackendoff, 2011, pp. 589–591 and footnote 1). However, in the last ten years, biolinguistics has been experiencing a turn for more emphasis on biology, which is well visible in articles such as *Prolegomena to a Future Science of Biolinguistics* and *Toward a computational framework for cognitive biology* by Tecumseh Fitch (2009, 2014) as well as *Biolinguistics: forays into human cognitive biology* by Cedric Boeckx (2013). Following those authors' designation, I call this 'new edition' of research on the biological foundations of language "cognitive biology of language" which is now (March 2019) also used as a name of the research laboratory led by Cedric Boeckx at the University of Barcelona, Spain.

In music research, the term "biomusicology" was introduced by Nils L. Wallin (1991) in his book *Biomusicology. Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music*. In *The Origins of Music* (S. Brown, Merker, & Wallin, 2000), biomusicology as defined as a discipline integrating evolutionary musicology, neuromusicology, and comparative musicology (i.e., cross-cultural studies investigating universals in music) was then taken up again. Almost at the same time, a conference entitled *The Biological Foundations of Music* (Zatorre & Peretz, 2001) brought scholars together who investigate theory, neuroscience, and evolution of music. While *The Origins of Music* focused more on evolutionary musicology, *The Biological Foundations of Music* rather concentrated on neuromusicology. In both cases, computational-representational theories of music investigating human cognitive capacity (Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee,

1984; Steedman, 1996; Sundberg & Lindblom, 1991) had earned less attention in light of biology.

Several turning points for more integrated approaches in biomusicology mainly happened in the last decade. In a conference proceeding of *The Neurosciences and Music II* (including, for example, Fitch, 2005; McDermott & Hauser, 2005a; Merker, 2005), a special issue on *The Nature of Music* edited by Isabelle Peretz in *Cognition* (including, for example, Fitch, 2006b; Jackendoff & Lerdahl, 2006) and a series of *Commentaries on Origins of Music* in *Music Perception* (including, for example, Fitch, 2006a; McDermott & Hauser, 2006; Patel, 2006), computational-representational theories of music, neuromusicology, and evolutionary musicology were ready to be integrated. Especially, through comparative language-music research, such integrative approaches were promoted. *Music, Language, and the Brain* by Aniruddh D. Patel (2008), *Brain and Music* by Stefan Koelsch (2012a), *Language and Music as Cognitive Systems* edited by Patrick Rebuschat, Martin Rohmeier, John A. Hawkins, and Ian Cross (2012), *Language, Music, and the Brain* edited by Michael A. Arbib (2013) are examples of such integrative challenges. The term "biomusicology" then resurfaced in an article *Four principles of bio-musicology* by Tecumseh Fitch (2015) as a part of Henkjan Honing's *The Origins of Musicality* research program (Honing, 2018; Honing, ten Cate, Peretz, & Trehub, 2015) – an integrative research program which can be called "biomusicology in a strong sense".

The current thesis is a part of those challenges to yield biolinguistics and biomusicology in the strong sense, and even more as a challenge to move toward human cognitive biology research discourse, with a particular emphasis on comparative approaches. Thus, the approach of the current thesis is called "comparative biomusicology", and the results of the thesis should show a way to move toward a research program which I call "comparative cognitive biology" by referring to W. Tecumseh Fitch's and Cedric Boeckx's research program, with an emphasis on comparative approaches. Especially, in the current thesis, I focus on language-music comparison because I argue that the nature of each cognitive system can be better understood in relation to other cognitive systems. As a comparative research strategy, I introduce *principled explanations* in terms of 1) syntax as a set of combinatorial principles generating hierarchically structured representations and 2) neural structures and operations as implementational principles. The idea of principled explanations in this thesis is not only influenced by that of Noam Chomsky (2005) investigating syntax, but also extends to more biological one inspired by Shepherd (2004) examining synaptic organization. In both cases, the central idea is explanation of complex phenomena in terms of basic principles applicable to properties in wide range of systems, regardless of cognitive systems or nervous systems. This

is at the heart of principled explanations which I introduce in the current thesis to promote comparative research and understand cognitive systems such as language and music.

2 "Syntax" or "structure building" for comparative language-music research?

Research on musical syntax got considerable attention through *generative syntax of tonal harmony* developed by Martin Rohrmeier (2007, 2011) by means of formal grammar. Rohrmeier and Pearce (2018, p. 475) reserved the term "syntax" for "approaches presenting a formal system characterizing the sequential structure of such building blocks [i.e., elements in a lexicon; RA]". That is, they characterize "syntax" in terms of formalization as a research method. On the one hand, I acknowledge the importance of formal approaches. On the other hand, I do not adopt this use of "syntax" in the current thesis because the goal of the current thesis is *not* to identify what aspects of language and music can be presented in a formal system. In addition, a definition which is too narrow and limited to a particular research method could hinder comparative approaches as phenomena which are investigated differently, but share biological substrates could not enter the analysis. Therefore, I rather use "syntax" in a broad sense applicable for a wide range of domains as a term referring to particular properties and functions. Moreover, while Rohrmeier and colleagues (2018; 2015) use "structure building" as an umbrella term to investigate structures in different domains, I regard this term as a particular part of "syntax".

In the current thesis, "syntax" is primarily used to refer to "a set of principles governing the hierarchical combination of discrete structural elements into lager units (Bickerton, 2009) and/or into sequences (Patel, 2008)" (Asano & Boeckx, 2015, p. 2). This characterization comprises processing theories, too, which was suggested by Rohrmeier and Pearce (2018) not to belong to the domain of syntax. If nothing else is stated in addition, "syntax" is used in this broad sense. Further, the current thesis suggests a core function of syntactic computation, i.e., a core task for syntactic computation to accomplish, as mapping between hierarchically structured representation and temporal sequence. Thus, I investigate syntax in the current thesis by identifying what principles are required and how those principles are applied to map hierarchical structure and temporal sequence in different domain such as syntax in language and music. Importantly, syntax in language includes different subcomponents such as narrow-sense syntax (i.e., combining words into phrases and sentences) and phonological syntax. Similarly, musical syntax contains subcomponents such as harmonic syntax and rhythmic syntax. Throughout the thesis, I will highlight those different subcomponents of syntax in language and music.

The following concepts are also relevant for the current approach to syntax. I use "syntactic computation" to refer to transforming inputs into outputs according to the principles that fulfill function of syntactic computation. For example, given words as inputs, syntactic computation is carried out on the basis of combinatorial principles and results in a hierarchical structure as its output. The term "parsing", then, refers to structural analysis during moment-by-moment integration of events in time. That is, parsing deals with online processing of a sequence. In both cases, the core function is mapping between hierarchical structure and temporal sequence, and a series of words should be transformed into a hierarchical structure. I use "(language) syntactic processing" [1] for phrasal syntactic and morpho-syntactic processing, and "music syntactic processing" for tonal-harmonic syntactic and rhythmic syntactic processing. The term "processing" includes both parsing and production, but the current thesis mainly focuses on parsing. Linguistic and musicological theories on syntax introduced in the current thesis are computational-representational theories because they attempt to clarify syntactic computation in detail, and identify representations which are generated on the basis of combinatorial principles and are processed in parsing.

3 Goals and structure of the thesis

The current thesis tackles the question "Why is music the way it is?" within a *comparative biomusicology* framework by focusing on musical syntax and its relation to syntax in language. As a comparative research strategy, I put forward principled explanations regarding cognitive systems as different instances of the same principles. The main goal is to provide a preliminary answer to this question in form of hypotheses about neurocognitive mechanisms, i.e., cognitive and neural processes, underlying a core function of syntactic computation in language and music, i.e., mapping between hierarchical structure and temporal sequence. The results, especially the insights into neural processes, should then contribute to establish *comparative cognitive biology* in future research. Thus, I see this thesis as an integrative challenge dealing with different methodological approaches (e.g., theoretical and empirical approaches), multiple cognitive systems (e.g., language and music), and a wide range of species (e.g., humans and non-human animals). Such an integrative challenge requires various perspectives and it is not possible to introduce all of them. However, I hope the current thesis to be an example for dealing with further integrative challenges and to contribute to a

[1] The term "syntactic processing" is used to refer to morpho-syntactic and phrasal syntactic processing, but I occasionally use "language syntactic processing" and "linguistic syntactic processing" in the same way if I discuss the relationship between language syntactic processing and music syntactic processing to avoid confusions.

realization of the original dream of cognitive science as a unified science of the mind (G. A. Miller, 2003, p. 144).

The thesis divides into four parts. The PART I *Principled explanations in comparative biomusicology* introduces a framework for integrative challenges, identifies emerging problems, and suggests principled explanations as research strategies to tackle those problems. Comparative biomusicology integrates different comparative approaches, biological frameworks, and levels of analysis in cognitive science. The problems which I will discuss in the PART I are those emerging in comparative research and in integrating different levels of analysis in cognitive science. PART II and III then implement the strategy which I call "principled explanations" from different perspectives. PART II is entitled *Top-down perspectives to syntax and its neurocognitive mechanisms* because the branches of comparative language-music research which I put together in this part mainly take top-down research strategies. They apply concepts of linguistic theories to musicological theories and those theoretical concepts to neuroscientific research. PART III is entitled *Rhythmic syntax: An integrative approach* because I attempt to unify both top-down and bottom-up perspectives in this part. In PART IV, I conclude the current thesis, discuss open questions, and suggest a possible way going toward comparative cognitive biology.

As PART III incorporates the central idea of the current thesis, I would like to introduce this part in more details in the next paragraph. First of all, it is important to note that, in general, I do not introduce any formal-mathematical theory of musical syntax paralleling Rohrmeier's *generative syntax of tonal harmony*. Rather, in PART III, I develop a computational-representational theory of rhythmic syntax which is not yet formally explicit, but already captures computational principles of rhythmic syntax. Till now, musical syntax was primarily associated with tonal-harmonic syntax. However, rhythmic syntax is another constituent part of musical syntax. The current thesis, thus, attempts to establish rhythmic syntax as a research field of musical syntax. I also elaborate rhythmic syntactic processing and its neural correlates on the basis of the developed theory. To identify neural correlates, ALE meta-analysis was conducted. This method recently gets increasing attention in music cognition research (e.g., Chauvigné, Gitau, & Brown, 2014; Janata & Parsons, 2013; LaCroix, Diaz, & Rogalsky, 2015) and turns out to be an important means for integrating theoretical and empirical research. Moreover, a large portion of PART III is dedicated to clarification of implementational principles and investigation of neurocognitive mechanisms in terms of cognitive and neural processes.

A comparative language-music research introduced in PART III focuses on the investigation of language and music as different instances of same principles. Especially, I attempt to take a step toward comparative cognitive biology by discussing neural structures and operations as implementational principles in details as well as introducing hypotheses how the same implementational principles give rise to different cognitive systems language and music. PART III notably puts mechanistic explanations at its heart. Thus, it does not put emphasis on the discussions about what is similar and different in language and music (for this line of research and discussions, see Asano & Boeckx, 2015; Jackendoff, 2009; Patel, 2008). Rather, I privileged discussions on cognitive and neural processes over the elaborations of other aspects. However, this does not mean that other aspects are irrelevant. It means that the current thesis leaves room for further integrative approaches.

PART I: Principled explanations in comparative biomusicology

The goal of this part is to specify the object of inquiry within the newly grounded approach called *comparative biomusicology*, identify problems to be solved, and develop research strategies. In particular, I introduce comparative and biological approaches to achieve a mechanistic explanation of music as a cognitive system (Chapter 4, p. 9), discuss emerging problems and first approaches (Chapter 5, p. 16), and suggest 'principled explanations' as research strategies to tackle those problems (Chapter 6, p. 30).

4 The research program and its framework

Comparative biomusicology is a research program which strives for an explication of the human capacity for music in relation to other cognitive systems such as language and action from a biological perspective. Its goal is to answer the question "Why is music the way it is?" by revealing the initial state and steady state of the cognitive system music. In general, as in biolinguistics (Boeckx & Grohmann, 2007), the following questions have to be investigated: What is the knowledge of music? How is that knowledge acquired? How is that knowledge put to use? How is that knowledge implemented in the brain? How did that knowledge emerge in the species? Therefore, biomusicology and biolinguistic strive for a biological foundation of the cognitive systems music and language from theoretical, psychological as well as neuroscientific, and evolutionary perspectives (Boeckx & Grohmann, 2007; S. Brown et al., 2000). Theoretical research such as linguistics and musicology focuses on computational-representational theory of music and language, psychology and neuroscience of language and music on cognitive and neural processes, and evolutionary biology on the phylogeny of language and music (Figure 4.1).

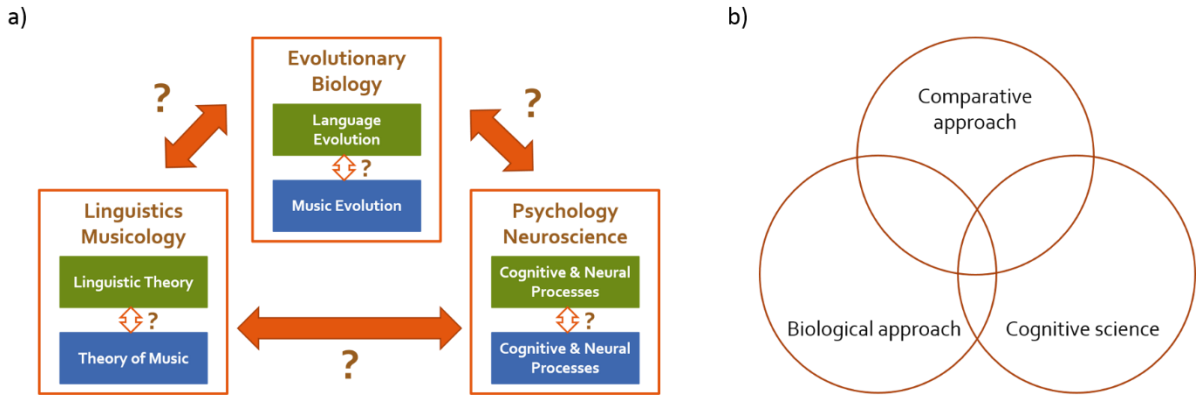


Figure 4.1 Comparative biomusicology as an integrated approach to cognitive systems.

Figure 4.1 shows diverse integrative challenges for comparative biomusicology. Figure 4.1a represents different disciplines such as linguistics, musicology, psychology, neuroscience, and evolutionary biology which comparative biomusicology includes. The red arrows indicate the challenges to relate those different disciplines to each other. The green boxes stand for language research and the blue ones for music research. The white arrows denote language-music comparative challenges. Figure 4.1b shows integrative challenges not in terms of disciplines, but of different approaches and frameworks which I illustrate below in more details.

4.1 Comparative approaches

Comparative biomusicology involves different comparative approaches such as 1) within-domain comparisons (e.g. cross-culture comparison), 2) between-domain comparisons (e.g. language, music, and action), 3) within-species comparisons (e.g. (developmental) disorders), as well as 4) between-species comparisons (e.g. non-human animals and humans). All those comparative approaches are crucial to specify and analyze the human capacity for music. Cross-cultural comparison can contribute to identifying constituent parts of the music capacity (Arom, 2000; S. Brown & Jordania, 2013; Nettl, 2000; Thompson & Balkwill, 2010). Comparisons of cognitive systems such as language and music as well as cross-disorder comparisons can further specify those constituent parts and approach to the quest for their modularity and domain-specificity (Jackendoff, 2009; Jackendoff & Lerdahl, 2006; Peretz, 2013; Peretz & Coltheart, 2003). Cross-species comparisons could shed light on the evolution of the constituent parts and the quest for human uniqueness (Fitch, 2006b, 2010a; Hauser, 2009; Hauser & McDermott, 2003; McDermott & Hauser, 2005b; Patel, 2006; Shettleworth, 2012).

A multi-component approach (Fitch, 2010a) identifying relevant constituents, levels, and their interactions is at the center of all comparative approaches (Asano & Boeckx, 2015; Fitch, 2015; Fritz et al., 2013; Honing et al., 2015). Language and music considered as cognitive systems form a mosaic and consist of multiple components with different evolutionary origins, which rely on separate neural and genetic mechanisms (Fitch, 2006b, 2010a). From a comparative language-music perspective, some components might be shared among the domains and based on the same evolutionary genesis, while others might be different and emerged independently in the course of evolution. From a comparative between-species perspective, some might be shared with other animals, while others might be unique to humans.

4.2 Biological frameworks

The most appropriate framework to achieve a biological foundation of music research is provided by Tinbergen's four questions and by distinguishing proximate and ultimate research questions (Bischof, 2008; Mayr, 1961; Tinbergen, 1963). Proximate causes underlie responses of an individual to immediate factors and are mechanical in nature, while ultimate causes account for evolution of particular genetic endowments (Mayr, 1961). By no means, those two causes should be confused as alternatives. However, proximate and ultimate causes are not fully independent of each other. Especially, development should be considered both proximately and ultimately causal. Thus, by taking evo-devo and niche construction into

account, Laland and colleagues (2011) argued for reciprocal causation allowing for interaction between two causes.

Tinbergen's four questions for explaining why animals behave the way they do were introduced shortly after Mayr's causal approach and include 1) "What are internal causal factors or mechanisms controlling behavior?" (causation / mechanism); 2) "How does behavior machinery change during development in an individual?" (ontogeny); 3) "How did behavioral traits evolve?" (evolution / phylogeny); and 4) "What is the survival value or the adaptive significance of behavior?" (function). Tinbergen's four questions are suggested as a framework for biolinguistics and biomusicology by several authors to investigate biological foundations of cognitive systems (e.g., Asano & Boeckx, 2015; Boeckx & Grohmann, 2007; Fitch, 2010a, 2015). Moreover, a current approach in biology suggests integrative solutions to the four questions instead of investigating them independently (Bateson & Laland, 2013).

One aspect introduced by Bateson and Laland (2013) plays a central role for the framework introduced in the Section 4.4 (p. 12) and thus should be mentioned here. As characterized above, Tinbergen's "function" is usually understood as ultimate question, i.e., adaptive significance, which is a consensus interpretation and should be maintained in the current thesis. However, Tinbergen's "function" can be also (or even more preferably) interpreted as proximate question i.e., current utility, "because it helps to emphasize how the current and original function of a characteristic can differ […]" (Bateson & Laland, 2013). This understanding is based on their evo-devo framework updating Mayr's and Tinbergen's biological frameworks. In the current thesis, Tinbergen's "function" stands for both adaptive significance and current utility.

I regard Mayr's distinction of proximate and ultimate causes, Tinbergen's four questions, and the updates by Laland and colleagues (2011) taking evo-devo and niche construction into account as biological frameworks. In the current thesis, I mainly focus on Tinbergen's biological framework (i.e., his four questions) and especially on the question of mechanism.

4.3 Levels of analysis in cognitive science

Marr (1982, pp. 19–29) introduced three levels at which an information processing device should be understood. The most abstract is the level of *computational theory* about *what* the device does, i.e. *what* is being computed, and *why*, i.e. *why* that particular computation is

performed and not another.[2] The latter can be formulated as *constraints* determining which computation is appropriate to use. The second level of analysis specifies *how* information is processed concerning *representation* of the input and output and an *algorithm* for the transformation. The third level deals with the way the process is to be realized physically, i.e. *hardware implementation*. Importantly, Marr emphasized that those three levels are coupled, but *only loosely* in the sense that all levels can influence each other while the explication of each level could involve issues independent of the other two. Especially, he advises caution in "making inferences from neurophysiological findings about the algorithms and representations being used, particularly until one has a clear idea about what information needs to be represented and what processes need to be implemented" (Marr, 1982, p. 26).

Among those three levels, Marr stresses the importance of computational theory from an information processing point of view even though algorithms and hardware are empirically more accessible: the information processing device cannot be specified without understanding the nature of the computational problems to be solved. In this sense, he calls Chomsky's transformational theory "a true computational theory" (Marr, 1982, p. 28). Moreover, Marr emphasized that those three levels of analysis are largely independent of each other. Thus, at that moment, the goal of any computational theories was solely identifying representations and computations which the information processing device deals with, without worrying about algorithms and hardware implementations.

4.4 A unified framework

Comparative biomusicology investigates the question "Why is music the way it is?" by comparative approaches within biological frameworks at different levels of analysis. Thus, the first task of this research program is to build a unified framework integrating all research strategies introduced above in an explicit way. In what follows, I aim at achieving such a conceptual synthesis gradually by beginning with relating Tinbergen's biological framework and Marr's levels of analysis, then integrating the multi-component approach to this unified framework, and finally discussing some emergent extensions.

There is an increasing tendency in modern research discourse to suggest a pluralistic explanatory framework integrating Tinbergen's four questions and Marr's three levels. For example, Poggio (2012), without referring to Tinbergen, suggested to extend Marr's three

[2] In the current thesis, I also use the term "computational-representational theory". This term covers Marr's computational theory and theory of representations.

levels with additional two levels on top: 1) learning and development, and 2) evolution, i.e., 1) proximate questions and 2) ultimate questions. That is, he claimed to revise Marr's three levels in light of biology. Krakauer and colleagues (2017) also see Tinbergen's four questions and Marr's three levels as important parts of pluralistic explanations. While Poggio (2012) as well as Krakauer and colleagues (2017) keep those two approaches separate, other researchers proposed a more intertwined view. Brase (2014) noted that Marr's levels should be revised to deal with *biological* information processing systems and Tinbergen's framework should be applied for investigating information processing underlying behaviors. He suggested that Marr's computational level in light of biology fits to ultimate causes, i.e., phylogeny and adaptive significance, and the algorithmic level to proximate causes, i.e., mechanism and ontogeny. Mobbs and colleagues (2018), by referring to Brase (2014), see Tinbergen's question about mechanism at Marr's algorithmic as well as implementational level and his question about function at Marr's computational level, while they regard ontogeny and phylogeny as the way how Marr's levels of analysis change over development and evolution.

Based on the above mentioned extensions and integrations, a framework for comparative biomusicology can be formulated as follows. First of all, the most straight-forward link between Tinbergen's questions and Marr's levels can be achieved by investigating mechanisms at the algorithmic and implementational level.[3] A *mechanism* is a system of different constituent parts performing unique operations and working in tandem to bring about a given phenomenon (Bechtel & Shagrir, 2015; Krakauer et al., 2017; Lobina, 2017). The mechanistic questions which can be asked at the algorithmic and implementational level are as follows: What kind / formats of representations do we need and how are they manipulated by the operations of the constituent parts to run a program, i.e., to process information?

Next, Bechtel and Shagrir (2015) suggested that the computational level identifies non-mechanistic aspects of a cognitive system, ideally in form of formal-mathematical descriptions, and clarifies the computed mathematical function in the physical environment (Bechtel & Shagrir, 2015). In other words, the computational level deals with formal-mathematical descriptions of problems which the mechanisms have to solve (what question) and functions for which the mechanisms are optimally designed (why question). Thus, although it does not seem to be straightforward, Tinbergen's question about function, i.e., not only current utility,

[3] I don't equate mechanisms with algorithms and hardware implementations as Marr and Poggio (1976) regard mechanisms as an independent forth level beside Marr's classical three levels. Bechtel & Shagrir (2015) and Lobina (2017) also regard the level of mechanisms as independent of the other three levels.

but also adaptive significance, can be linked to the computational level of the analysis, especially to the why question.[4]

Finally, ontogeny and phylogeny can be regarded as change of information processing over development and evolution. This is similar to the interpretation of Mobbs and colleagues (2018): Marr's levels are levels of analysis and thus change of information processing should be analyzed at all three levels. Therefore, the questions can be formulated as following: "What is computed and why is that particular computation performed at a certain developmental or phylogenetic stage?" (computational level); "How is information processed at a certain developmental or phylogenetic stage?" (algorithmic level); "How is the process realized physically at a certain developmental or phylogenetic stage?" (implementational level); and also "How do computations, representations and algorithms, and implementations change over development and evolution?".

The unified framework can be regarded as a *comparative biological information processing framework* (Figure 4.2). In this framework, mechanisms play a mediator role for biology and information processing. This does not mean that we have to reduce our research to mechanisms. However, this means that the investigation of biological information processing mechanisms is at the center of comparative biomusicology. Explanations at the computational level and answers to questions about function, as a formal-mathematical theory in the optimal case, provide top-down constraints on the mechanisms, in the way that they determine what problems the mechanisms have to solve and why. Then, the mechanisms can be analyzed at the algorithmic and implementational level. The changes of mechanisms and their goals over development and evolution can be also investigated. In addition, different constituent parts of the mechanisms can be identified by means of a multi-component approach, which opens the current framework for comparative approaches. Because mechanisms can be investigated at the very bottom level (e.g., molecular level) as well as the 'higher' level (e.g., behavior), they offer rich comparative options (Bateson & Laland, 2013).

[4] In Figure 4.2, I did not explicitly relate Tinbergen's question of function and Marr's computational level as it is still a matter of debate whether and how they relate to each other. However, research at the computational level investigating why a particular computation is performed could be related to research on the current utility and the adaptive significance.

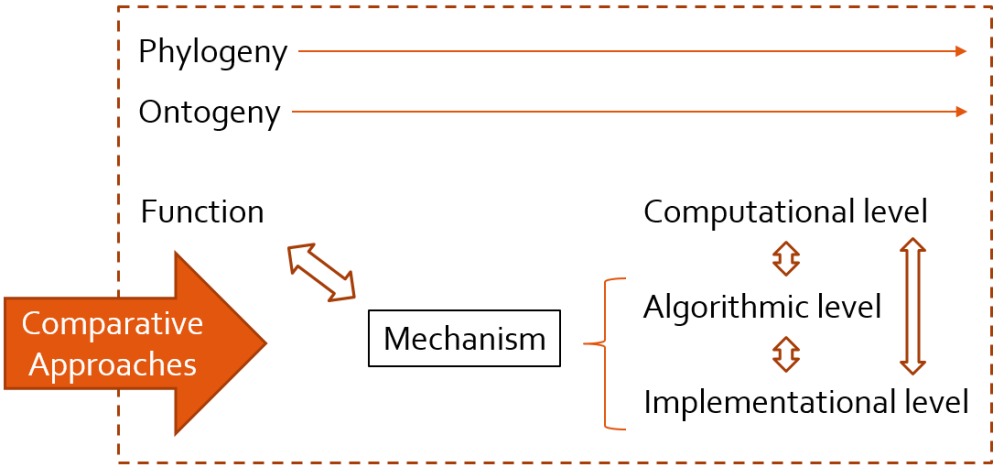


Figure 4.2 Comparative biological information processing framework.

In addition, the current comparative biological information processing framework should be extended to integrate cultural and social aspects as suggested by several authors (e.g., Fitch, 2010a, 2015; Laland et al., 2011). First of all, the framework introduced above (see also Figure 4.2) keeps the strict separation of proximate and ultimate questions in terms of ontogeny and phylogeny. Thus, there is no chance for those two types of mechanism change to influence each other. However, such a proximate-ultimate dichotomy was claimed as a conceptual barrier and a reciprocal causation approach was suggested as an alternative way (Laland et al., 2011).

> During development, features of the trait cause changes in both gene expression and environment, which feed back to the developmental process, resulting in a different trait in the adult and modifications of both developmental and selective environments. (Laland et al., 2011, p. 1514)

In a similar line, Fitch (2010a, pp. 33–34) introduced *glossogeny* as a distinctive, intermediate level concerning culturally transmitted change. Based on such extensions, I propose to think of an additional question in terms of social interaction and cognition (e.g., Han et al., 2013; Iacoboni, 2009; Koelsch, 2010; Laland et al., 2011; Tomasello, 1999; Vogeley & Roepstorff, 2009; Ward, 2012). Taken this challenge into account, comparative biomusicology aims to ground and explain social, cultural and historical musical phenomena within a biological framework without neglecting their relevance. How to address social interaction and social cognition in comparative biomusicology remains an open but important topic. For now, this additional question is still underspecified and needs further refinement, but further

considerations of this question will be very fruitful for the development of comparative biomusicology. In the PART IV (Section 17.1, p. 171), I will briefly come back to this issue.

5 Methodological problems and first approaches

5.1 Problem of contrastive comparison

In the Chapter 4 (p. 9), I introduced a comparative biological information processing framework and suggested a mechanistic approach, especially a multi-component approach. However, contrasting domain- and species-specific constituents also doesn't lead to true comparative research. Therefore, current comparative approaches agree upon the importance of such a divide-and-concur strategy, but, at the same time, are worrying about what the right level of comparison is. In what follows, I discuss this issue from two perspectives: a comparative approach to language and music and a comparative cognition approach.

5.1.1 The comparative approach to language and music

To avoid a contrastive approach, it is a good starting point to clarify the constituent parts of language and music at Marr's three levels of analysis. In comparative language-music research, Fritz et al. (2013) suggested such an approach by largely focusing on the algorithmic and implementional level.[5] For Fritz et al. (2013, p. 420), "perception and active performance of music (including song and dance) as well as language comprehension and production" are placed at the level of computation. However, this statement neither explains *what* is computed nor *why*. As input and output representations, they list up several primitives (see Table 5.1). Those representations are suggested as domain-specific. Moreover, concerning *algorithms* for the transformation, several processes are listed (see Table 5.1), which were claimed as domain-general. Fritz and colleagues (2013) hypothesized that domain-specific representations can be processed by domain-general algorithms, but did not elaborate this discussion. The hypothesis is stated by Fritz et al. (2013, p. 425) as following:

> In contrast to the representational inventories, we hypothesize that many of the algorithms/operations that have such primitives as their inputs are, by and large, domain-general or, at least, will prove to combine generic algorithms in domain-specific ways. One way to conceptualize this is to imagine different

[5] Fritz and colleagues (2013) use the term "algorithm" to refer to cognitive processes and they do not introduce any concrete algorithm. This is also the case in the current thesis because I investigate algorithm in terms of cognitive processes for (psychological) parsing. See also PART IV, Section 17.3, p. 176 for discussions.

invocations of the same neural circuitry; that is, "copies" of the same circuitry, but which operate on input representations of different types that are domain specific.

Table 5.1 Elementary parts list (preliminary).[6]

Marr's levels		Language	Music
Computational			
Algorithmic	Representation	(articulatory) Feature; Phoneme; Syllable; Morpheme; Phrase (e.g., noun and verb phrase); Clause; Sentence; Discourse and narrative	Note (timbre and pitch); Pitch interval (dissonance and consonance); Octave-based pitch scale; Pitch hierarchy (tonality); Discrete time interval; Beat; Meter; Motif / theme; Melody / satz; Piece
	Algorithm	Constructing spatiotemporal objects (streams, gestures); Extracting relative pitch; Extracting relative time; Discretization; Sequencing, concatenation, ordering; Grouping, constituency, hierarchy; Establishing relationships: local or long distance; Coordinate transformations; Prediction;	

[6] This elementary parts list was suggested by Fritz et al. (2013, p. 420). The labels of Marr's levels were called differently by Fritz et al. (2013), but have been changed here to avoid terminological confusion. I interpret their "representational computational" and "algorithmic computational" as belonging to the "algorithmic level", and called the former "representation" and the latter "algorithm". In addition, I removed "domain-general" and "domain-specific" from the list because it is difficult to classify the levels according to those labels (see the discussion in the text.

		Synchronization, entrainment, turn-taking; Concurrent processing over different levels
Implementational		Generic forms of circuitry; General learning rules which can adapt circuits to serve one or both domains

Throughout the current thesis, in concert with Fritz et al. (2013), I argue that contrasting language and music domain-specific representation is not a fruitful comparative approach. Especially, adapting specific concepts developed in linguistics to music might not be as fruitful as expected or, even worse, harmful for comparison (Asano & Boeckx, 2015). However, in contrast to Fritz et al. (2013), I suggest that even at the level of representation, there is an important similarity between language and music, namely headed hierarchical structures. I will even make a stronger claim that comparative research on language and music should take a *bottom-up approach* investigating cognitive systems in terms of *a set of common principles*. Of course, such a comparative approach should also explain differences. In explaining similarities and differences within a comparative framework, I take the hypothesis on domain-general algorithms operating on domain-specific representations seriously and elaborate this idea from theoretical as well as from empirical perspectives to achieve a comparative approach without being contrastive. In any case, such an *elementary parts list* ranging in all levels is an important and necessary starting point.[7]

How does the elementary parts list of the current thesis look like? As mentioned above, Fritz and colleagues (2013) did not characterize language and music at the computational level explicitly. Thus, I begin with updating this level by identifying what is computed and why. At the computational level, language is characterized in terms of mapping sound and meaning (Chomsky, 2010). For music, linking sound and affect is the computational problem to be solved (Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1983). At the algorithmic level, linguistic and musical parsers should be investigated. As for music, two central aspects of music processing should be explored: *Tonal encoding of pitch*, i.e., computing pitch in relation to the tonal center, and *beat-based encoding of rhythm*, i.e., computing rhythm in relation to a periodic internal beat. These components were suggested to be potential innate specifications

[7] An elementary parts list based on results of the current thesis is presented and discussed in the PART IV.

for music (Honing & Ploeger, 2012). Then, the implementational level should identify how the parsers are realized in terms of neural structures or circuits and operations carried out by them.

5.1.2 The comparative cognition approach

Determining constituents for comparative approaches between species based on complex mental abilities of humans might also lead to inconclusive contrasting. De Waal and Ferrari (2010) emphasized this problem and suggested a bottom-up perspective to between-species comparative approaches attempting an exploration of basic processes. They put the following questions at the center (De Waal & Ferrari, 2010, p. 201): How does cognitive capacity X actually work? "What are the necessary ingredients of X and how did these evolve?". Those questions emphasize two of Tinbergen's four questions: mechanism and phylogeny. By discussing mental capacities memory and planning, imitation as well as prosocial behavior and empathy, they challenged the claim of human uniqueness and put forward an alternative view (De Waal & Ferrari, 2010, p. 205): "*De novo* appearance of cognitive capacities are apparently as unlikely as *de novo* anatomical features". If we take the questions of mechanism and phylogeny as well as the possibility of evolutionary continuity in complex mental abilities seriously, what might be the constituents for between-species comparative approach to language and music?

First of all, a multi-component approach which avoids technical theoretical terms from linguistics or musicology, but uses "song" and "dance" (shared with some bird species) as well as "drumming" (shared with some non-human primate species) to identify the behavioral domains would be a good starting point for between-species comparison (Fitch, 2006b, 2015). Moreover, other researchers suggested mammalian calls, i.e. acoustic communication, as promising candidate (Ackermann, Hage, & Ziegler, 2014; Fitch & Zuberbühler, 2013; Jackendoff, 2009; Rauschecker, 2013), while others put gestural communication forward (Arbib, 2011). As for underlying capacities relevant to the discussion of language and music evolution, vocal learning or vocal control (Fitch, 2011; Patel, 2006), motor control (Fujita, 2016; Lieberman, 2010), social learning (De Waal & Ferrari, 2010; Tomasello, 1996; Zentall, 1996), and imitation (Arbib, 2011) can be listed. Regarding all those candidate domains and capacities, one of the central questions in comparative language and music evolution research is what makes the human brain differ from that of other animals in being language- or music-ready, i.e. 'equipping' for language and music, and how it got that way.

In comparing different species, two distinctive classes of shared traits are identified: *homologies* which are derived from a shared trait existing in the common ancestor and

analogies which evolved independently in multiple lineages (Fitch, 2017). In addition to those two, there is *deep homology* which arose independently in multiple distant lineages at a phenotypic level, but build on the genetic and developmental mechanisms present in the common ancestor (Fitch, 2017). For example, concerning the components mentioned above (Section 5.1, p. 16), drumming is a homological trait which likely existed in the last common ancestor (LCA) of humans and chimpanzees, while song is an analogous trait, which emerged by convergent evolution in distant species, or is possibly a case of deep homology (e.g., Foxp2 gene). Comparative cognition research, then, investigates similarities and differences of cognitive and neural mechanisms between those species.

While some researchers question the usefulness of such an approach in its current form (e.g., Hauser et al., 2014), others regard it as one of promising methods to investigate the evolution of language empirically (e.g., Fitch, 2017). Importantly, any animal model can be a model of human language. Thus, each animal model can potentially inform us about some subcomponent of language. For example, song bird models contribute to the investigation of vocal learning, but might not be informative to elucidate the evolution of semantics. Moreover, in light of comparative cognition, "correct inferences about the relationship between cognitive or brain processes in humans and those in nonhuman animals depend on a detailed appreciation of the biology of 'animal models'" (Shettleworth, 2010, p. 16). Therefore, an important question in choosing an animal model is whether and how it can contribute to mechanistic understanding of a subcomponent of human language or music in terms of cognitive and neural processes.

In approaching this question, a corticocentric myopic view contrasting the cortex implementing 'higher' functions and subcortical structures implementing 'lower' functions should be avoided: cortical and subcortical structures are relating to each other to implement the 'higher' functions (Parvizi, 2009). Thus, integration of subcortical structures is necessary for a bottom-up comparative approach. As suggested by De Waal and Ferrari (2010, p. 202), "[e]very species, including our own, comes with an enormous set of evolutionarily ancient components of cognition that we need to better understand before we can reasonably focus on what makes the cognition of each species special". They further ask whether "[…] cognitive specializations [are] due to new capacities or rather to new combinations of old ones" (De Waal & Ferrari, 2010, p. 202).

Throughout the current thesis, to achieve a non-contrastive approach, I hold a view against corticocentrism and highlight the role of evolutionarily ancient components in cognition and the possibility of language and music having emerged as new combinations of

old capacities. In particular, I emphasize key roles of the basal ganglia in cognition and suggest that they belong to one of ancient components combined in new ways within the cortico-basal ganglia-thalamocortical circuits, yielding one aspect of language- and music-ready brain. However, this does not mean that I ignore the cortex. Rather, by investigating the cortico-basal ganglia-thalamocortical circuits, I attempt to integrate approaches investigating both cortical and subcortical structures.

5.2 The problem of explanatory gaps

First of all, it is worth noting that the problem of explanatory gaps comprises two separate problems: a) How is mapping from theoretical concepts of music theory and linguistic theory onto hypothetical constructs of psychology established to get psychologically relevant and testable concepts?; b) How to deal with the explanatory gap between mind and brain in comparative biomusicology? In language cognition research, those problems have been discussed repeatedly. Question a) relates to the debate on how to deal with the competence-performance distinction and question b) to the discussion about how to relate mind and brain. In what follows, first, I briefly summarize the state of the art regarding question a) and b), and first approaches suggested in language cognition research. Then, I propose how comparative biomusicology might profit from those debates and approaches.

5.2.1 Going beyond the competence-performance dichotomy

After the cognitive revolution in the 1960's, linguistic theory and psychology as disciplines of cognitive science have been concerned with the internal mechanisms associating sound and meaning in a particular way. Chomsky puts emphasis on the internal mechanisms by differentiating *competence* – the capacity / knowledge of the ideal speaker-hearers to map sounds and meanings strictly in concert with the rules of their language – and *performance* – the actual use of this capacity / knowledge in a certain situation (Chomsky, 1965, 1968). As he states it (Chomsky, 1965, pp. 3–4):

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. [...] To study actual linguistic performance, we must consider the interaction of

> a variety of factors, of which the underlying competence of the speaker-hearer is only one. [...]
>
> We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence.

Based on this agenda of linguistic theory introduced by Chomsky, generative linguistics has focused on the investigation of competence in terms of computations and representations generating language. Performance is, then, an interaction effect of the competence with a variety of factors such as memory limitations, distractions, shifts of attention and interest, and errors (Chomsky, 1968).

Moreover, Chomsky (at least in 1968) characterizes a *grammar* (*G*) as a competence model for linguistic inquiry and, in parallel, a *perceptual model* (*PM*) as a performance model for psychological research. On one hand, he points out the parallel between *G* and *PM* in the sense that both are about internal mechanisms associating sounds and meanings. On the other hand, he clearly distinguishes those two levels of investigations. In his words (Chomsky, 1968, p. 104):

> Both *G* and *PM* relate sound and meaning; but *PM* makes use of much information beyond the intrinsic sound-meaning association determined by the grammar *G*, and it operates under constraints of memory, time, and organization of perceptual strategies that are not matters of grammar. Correspondingly, although we may describe the grammar *G* as a system of processes and rules that apply in a certain order to relate sound and meaning, we are not entitled to take this as a description of the successive acts of a performance model such as *PM* - in fact, it would be quite absurd to do so.

This passage, as I understood, emphasizes that the derivation, i.e. stepwise generation of a syntactic representation, does not conform to online processing steps in performance. This is one of the reasons why linguistic theory and psycholinguistic research went in different directions. For decades, this strict distinction between competence and performance was maintained. However, the research on the biological foundations of language can be achieved

only if both linguistic competence theory and empirical research areas investigating performance such as psychology and neuroscience are integrated.

Beyond differences in theoretical positions, researchers recently agreed upon the view that differentiating competence and performance does not mean that linguistic theory is independent of any empirical data, nor psycholinguistics as well as cognitive neuroscience of language work regardless of linguistic theory (Boeckx, 2010; Ferreira, 2005; Jackendoff, 2002; Marantz, 2005; Myachykov, Tomlin, & Posner, 2005). Two major strategies for constructing mapping hypotheses between competence and performance, i.e. tackling the mapping problem a), dominate current language cognition research so far (Figure 5.1).

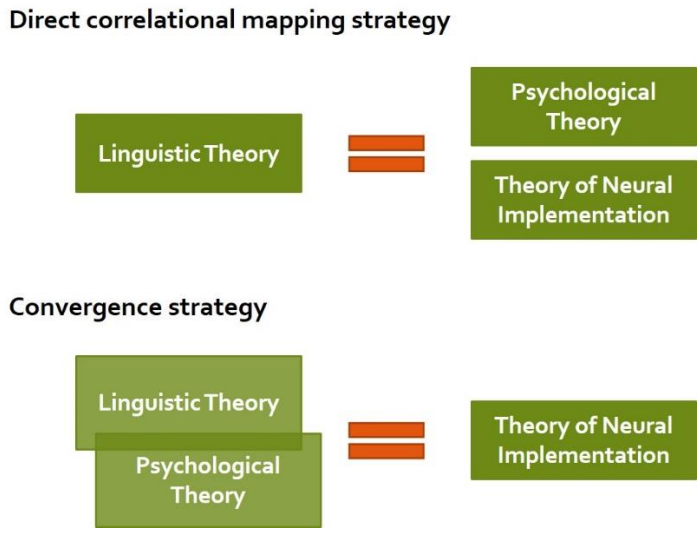


Figure 5.1 Two major strategies for constructing mapping hypotheses between competence and performance

The first one is the *direct correlational mapping strategy* connecting linguistic theory directly to the measures of psycholinguistics and cognitive neuroscience. For example, Marantz (2005) discussed how generative linguistic theory acts as a theory of language within cognitive neuroscience and suggested that hypotheses within all generative theories can be empirically tested by correlating representational and computational complexity with experimental dependent variables. Given the ideal situation in which all other variables are equal, the more complex a representation, the longer a subject requires to perform a task and the more activity in the brain should be observed. Therefore, he claims that "the categories and operations of generative grammar are hypotheses about the representations and computations in the minds and brains of speakers" (Marantz, 2005, p. 440).

The other type is the *convergence strategy* by developing competence theory with the principles resembling to the principles that the language user actually employs in performance (Jackendoff, 2002). Thus, this approach posits a competence theory with a close relation to a theory of processing and attempts to integrate concepts from processing theory such as working memory to the competence theory. Jackendoff's approach is based on the *soft* competence-performance distinction and advocates the integration of linguistic theory into a broader psychological context. Therefore, Jackendoff (2002, p. 34) views the investigation of the human language capacity as follows:

> Theory of competence: the functional characterization of the "data structures" stored and assembled in the f-mind [functional mind; RA] in the course of language use.
> Theory of performance: the functional characterization of the use of these data structures in the course of language perception and production.
> Theory of neural instantiation: how the data structures and the processes that store and assemble them are realized in the brain.

According to those definitions, especially that of the theory of competence, it can be said that Jackendoff's enterprise is geared to usage-based approach aiming at the explanation of online language processing. Especially, his theory of language is an integration of the generative grammar (computational theory) and Baddeley's theory of working memory (psychological theory) (Baddeley, 2010).[8]

What might be an appropriate strategy to deal with problem a) and b) in comparative biomusicology? A possible answer to question a) might be tackled by establishing the relationship between music theoretical concepts and hypothetical constructs of psychology. A promising starting point is working out the relationship between theoretical investigations of musical structure in terms of grouping, meter, and tonal hierarchies (Jackendoff & Lerdahl, 2006; Lerdahl, 2013; Lerdahl & Jackendoff, 1983) and psychological approaches studying musical expectancy building based on processing complex hierarchical structure – musical event hierarchy including both harmonic and rhythmic aspects – in time within certain context (Patel & Morgan, 2017; Rohrmeier & Koelsch, 2012; Tillmann, 2012).

[8] For recent reviews including updates of the model, see, for example, Baddeley (2010, 2012)

5.2.2 Approaching the mind/brain interface problems

Poeppel and Embick (2005, pp. 1–5) pointed out two *interface problems* in studying cognition at the linguistic-neuroscience interface. The first one is called the *Granularity Mismatch Problem* (*GMP*), which means that the elemental concepts of linguistics and neuroscience (neurobiology as well as cognitive neuroscience) do not match in their 'conceptual granularity'. They particularly suggested that linguistic concepts possess more fine-grained distinctions than those in neuroscience. The second problem is called the *Ontological Incommensurability Problem* (*OIP*): the constituent elements of linguistic theory cannot be put on a level with the fundamental biological units in neuroscience. A solution of those problems, according to Poeppel and Embick (2005, p. 5), is "spelling out the ontologies and processes in computational terms that are at the appropriate level of abstraction (i.e. can be performed by specific neural populations) such that explicit interdisciplinary linking hypotheses can be formulated".

The most parsimonious approach concerning this second problem is to assume that it will not be possible to close that gap. Although it is important to keep the difficulty in mind, modern cognitive science can only make progress by approaching this problem. One way to tackle problem b) is *explanatory neurolinguistics* which aims at establishing explanatory relationship between the computational-representational (CR)[9] theory of language and the neurobiological (NB) domain (Embick & Poeppel, 2015). Within this framework, they suggest to investigate the explanatory connections between CR and NB on the basis of Marr's levels. There are three possibilities: Type I Computational/Hardware, Type II Algorithm/Hardware, and Type III Computation/Algorithm. Type III implicates mappings between theoretical linguistics and psycholinguistics theory and is relating to the issue discussed above concerning mapping problem a). Although Type I explanation, i.e. the explanation of why certain NB structures computes particular computations and not the others, is the ultimate goal of the explanatory neurolinguistics, it is still unclear at present whether this mapping can be yielded. Embick and Poeppel (2015) suggest that Type II explanations, i.e. explaining why certain NB objects are optimally designed to implement particular procedures, can be yielded in a more direct manner than Type I explanations. To their endeavor, moreover, the following two questions are central (Embick & Poeppel, 2015, p. 363):

[9] The computational-representational theory subsumes Marr's computational and algorithmic levels (Embick & Poeppel, 2015).

Specialisation question 1: Are there particular levels of NB organization that are to be privileged as candidates for CR specialisation?

Specialisation question 2: Are there particular parts of the CR theory that are more likely to be candidates for explanatory neurolinguistic explanation than others?

Concerning the first question, they suggest the level of the neural circuit as an appropriate level to investigate the relationship between CR and NB. The second question cannot be answered at the moment, especially given the difficulty to establish Type I explanations.

In approaching the Type II explanation, the relationship between cognitive processes and brain activities should be carefully evaluated. A logically valid kind of inference drawn from neuroimaging data usually has the following form (Poldrack, 2006, p. 59): "if cognitive process X is engaged, then brain area Z is active". From this form, it is possible to deduce a conclusion in two logically valid ways: "cognitive process X is engaged, thus brain area Z is active" (*modus ponens*) and "brain area Z is not active, thus cognitive process X is not engaged" (*modus tollens*). However, Poldrack (2006) indicates that reasoning made in experimental neuroscience articles is based on a 'reverse inference' which is basically a logically invalid form, i.e. the logical fallacy of affirming the consequent: "brain area Z is active, thus cognitive process X is engaged". Logically invalid reasoning is per se not a bad reasoning and "reverse inference might be useful in the discovery of interesting new facts about the underlying mechanisms" (Poldrack, 2006, p. 60), but it should be used with caution because such an inference does not necessarily lead to a true conclusion. Poldrack (2006) suggests to restate the inference in probabilistic terms by using one form of Bayes' theorem:[10]

(5-1) $$P(COG_X|ACT_Z) = \frac{P(ACT_Z|COG_X)\,P(COG_X)}{P(ACT_Z)}$$

where

(5-2) $$P(ACT_Z) = P(ACT_Z|COG_X)\,P(COG_X) + P(ACT_Z|\neg COG_X)\,P(\neg COG_X)$$

[10] COG_X refers to "the engagement of cognitive process X" and ACT_Z to "activation in region Z" (Poldrack, 2006, p. 60). According to Bayes' theorem, the conditional probability of COG_X given ACT_Z (known as the posterior probability) can be determined in terms of three components: the conditional probability of ACT_Z given COG_X––the "likelihood"—, i.e., $P(ACT_Z|COG_X)$, the 'stand-alone' probability of COG_X without any context (known as the prior probability), i.e., $P(COG_X)$, and the conditional probability of ACT_Z in the absence of COG_X, i.e., $P(ACT_Z|\neg COG_X)$. The 'stand-alone' probability of ACT_Z without any context, i.e., $P(ACT_Z)$, (known as the base rate of ACT_Z) can be calculated by means of these three components as stated in the second equation (2). For a short, precise characterization of this theorem, see, for example, chap. 7.2 in Tanimoto (1990, pp. 330–335).

This means, we regard degrees of belief in a reverse inference as probability values, i.e., $P(COG_X|ACT_Z)$.

The reason why this probabilistic inference is important can be clarified on the basis of Figure 5.2 (Poldrack, 2006, p. 61) which illustrates the relationship between experimental manipulation through tasks, (unobservable) cognitive processes, and (observable) measure in a probabilistic graph, i.e., Bayesian network. This figure shows that the degree of belief in a reverse inference (from observable measures to unobservable processes), i.e., $P(COG_X|ACT_Z)$, depends on "the prior belief in the engagement of cognitive process X given the task manipulation", i.e., $P(COG_X|TASK_Y)$, and "the selectivity of the neural response (i.e. the ratio of process-specific activation to the overall likelihood of activation in that area across all tasks)", i.e., $P(ACT_Z|COG_X)$ (Poldrack, 2006, p. 60). The first aspect relates to the question "How likely does $TASK_Y$ cause COG_X?" and the second aspect to "How probably does COG_X cause ACT_Z?". If the relationship between task and cognitive process is not clear, $P(COG_X|TASK_Y)$ might be at the chance level (e.g. 0.5). If not only a particular cognitive process 1, but also other cognitive processes 2, 3, 4, and so on cause the activity in the same brain region or the same behavioral data, the selectivity of $P(ACT_Z|COG_X)$ might be low. Thus, optimizing those two aspects leads to the improvement of reverse inferences.

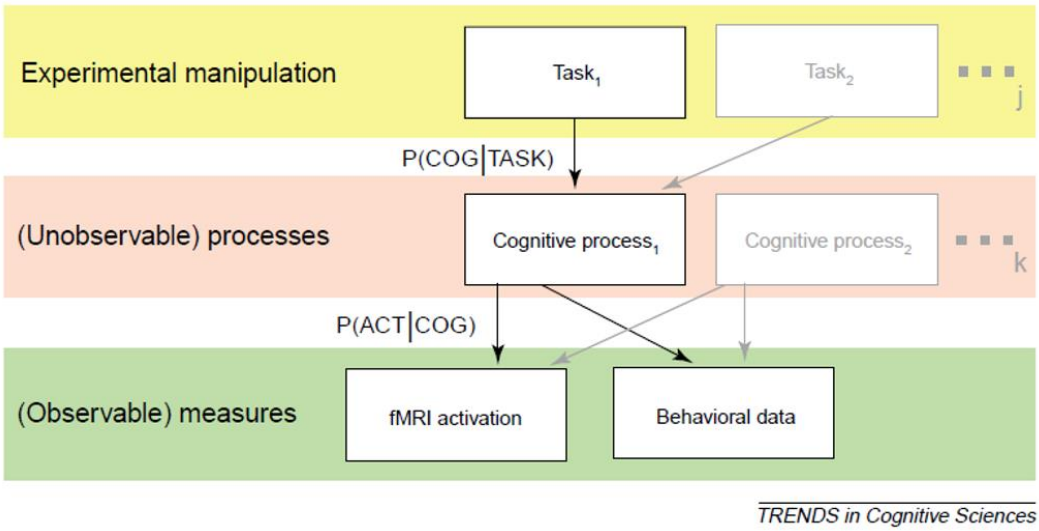


Figure 5.2 Relationship between experimental manipulation through tasks, (unobservable) cognitive processes, and (observable) measure in a probabilistic graph.[11] This figure is adopted from Poldrack (2006, p. 61), *Trends in Cognitive Sciences*, 10 (2): 59-63, with permission from Elsevier.

[11] To note that $P(COG_X)$ in the above mentioned equations are represented hier as the probability of COG_X given $TASK_Y$, i.e. $P(COG_X|TASK_Y)$.

Neurocognitive modelling might be one of strategy optimizing $P(COG_X|TASK_Y)$. Jacobs and Hofmann (2013) claim that the more precise the specification of cognitive processes or implementation in a simulation model, the higher $P(COG_X|TASK_Y)$ becomes. Thus, they set the goal of neurocognitive modeling as identifying core processes which underlie effort (measured in terms of brain activities or behavioral data) in all experimental paradigms inducing a particular cognitive process such as word recognition. They also suggest that a successful neurocognitive model should optimize $P(ACT_Z|COG_X)$ by specifying the function of a particular neural structure in a certain cognitive function as precisely as possible. Thus, building an effective neurocognitive model requires systematic relationships between cognitive function and neural structure, i.e. a detailed cognitive or functional ontology[12] (Jacobs & Hofmann, 2013; Poldrack, 2006; Price & Friston, 2005).

There are several approaches to specify functions of neural structures. The first one suggested by different authors (e.g., Jacobs & Hofmann, 2013; Poldrack, 2006; Price & Friston, 2005) is the network approach to the brain (Wilkins, 2018 for neuromusicology), especially by using effective connectivity (also called functional connectivity) analysis. This means that the structure-function mappings are performed not at the level of single anatomical structures, but at the network level by analyzing sets of functionally or effectively interconnected and co-activated regions. According to this approach, one region activated over different cognitive functions might show selectivity to a particular function in terms of the other co-activating regions. Another one is a cross-disorder approach to explore the question "What neural structure is *necessary* for a particular cognitive function and what is just involved in it?". A breakdown of cognitive functions through brain lesions provides strong evidences for functional selectivity. Therefore, an intertwined research strategy of brain imaging and cross-disorder approach is a key to construct functional ontology (see Figure 5.3) (Price & Friston, 2005, p. 271).

[12] Price and Friston (2005, p. 269) suggested that a cognitive or functional ontology should integrate structural (i.e., neuroanatomical) and functional (i.e., cognitive function related) information, so that a given cognitive function predicts the brain structures engaged and vice versa. As shown in Figure 5.3, cognitive functions are hierarchically organized into different granularities of sub-processes. They further claim that relationships between brain activations should be reflected in the cognitive ontology in terms of functional interactions inferred by co-activations in different brain regions rather than neuroanatomical connections. Such systematic structure-function relationships should be explicitly determined and implemented by means of data base systems, for example.

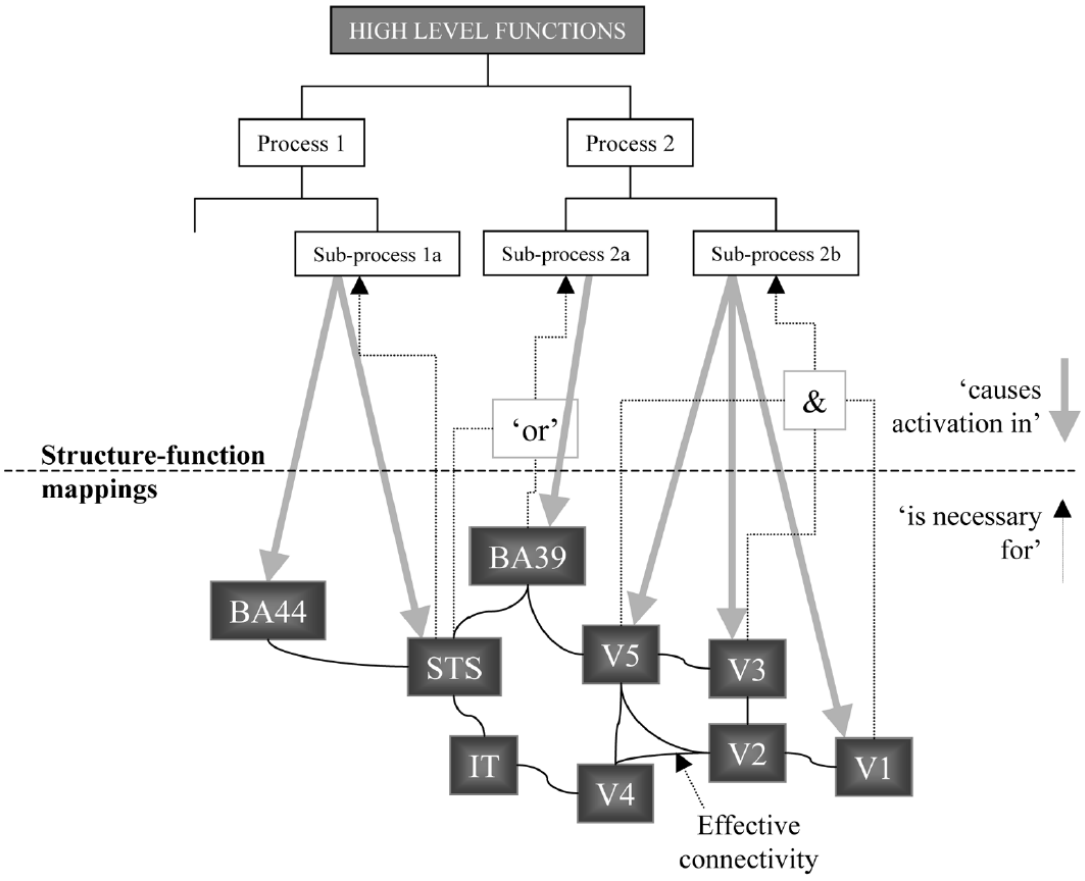


Figure 5.3 Structure-function mappings. This figure is adapted from Price and Friston (2005, p. 271), *Cognitive Neuropsychology*, 22 (3-4): 262-275, with permission by Taylor & Francis.

If we regard language and music as neurocognitive systems, the object of the investigation is (unobservable) neurocognitive processes. Therefore, the approaches suggested above are necessary. Music cognition research indeed makes use of reverse inference. A prominent example is represented in a sentence like "Because Broca's region is activated, therefore, music also engages hierarchical structure processing". This sentence seems to be based on the following reverse inference: Premise 1 "If hierarchical structure processing is engaged, then the Broca's region is active"; Premise 2 "The Broca's region is active"; Conclusion "Hierarchical structure processing is engaged". In such a case, however, not only the tasks, but also the stimuli should be considered to optimize the selectivity of an unobservable cognitive process given a particular experimental manipulation. This is one missing component in the approach of Poldrack (2006). That is, we have to consider $P(COG_X|TASK_Y \text{ and } STIMULUS_R)$.

Moreover, Price and Friston (2005) discuss neural correlates of cognitive processes in terms of neural structures and not neural processes. However, comparative biomusicology

attempts to explain biological information processing mechanisms in terms of cognitive and neural processes. Thus, not only neural structures, but also operations carried out by particular neural structures should be identified and integrated in such an approach. The reverse inference at the edge of $P(ACT_Z|COG_X)$ can be also optimized by specifying the neural processes which can be implemented by particular neural operations implemented in particular regions. The candidate neural processes could then contribute to narrow down conceivable cognitive processes among many alternatives.

6 Principled explanations as research strategies

Comparative biomusicology investigates the question "Why is music the way it is?" in relation to other cognitive systems (e.g., language and action) and other species (e.g., non-human primates and song birds) within a unified comparative biological information processing framework in which a mechanistic approach concerning cognitive and neural processes plays a central role. Figure 6.1 summarizes the issues discussed in the PART I and serves as the basis of the current thesis. The vertical double arrow between computational and algorithmic levels indicates the competence-performance or theory-psychology mapping problem. In the current thesis, I discuss this relationship in terms of mapping hierarchical structure and temporal sequence as a core function of syntactic computation. In this way, I aim at providing Type III explanation introduced by Embick and Poeppel (2015). The *mind/brain interface problem* for the Type II explanation should be investigated in terms of the relationship between cognitive and neural processes as a question about mechanism. In the current thesis, thus, I approach the Type II explanation by discussing to what extent cognitive and neural processes can be related. In Figure 6.1 this challenge is indicated by the equal sign between cognitive and neural processes with a question mark.

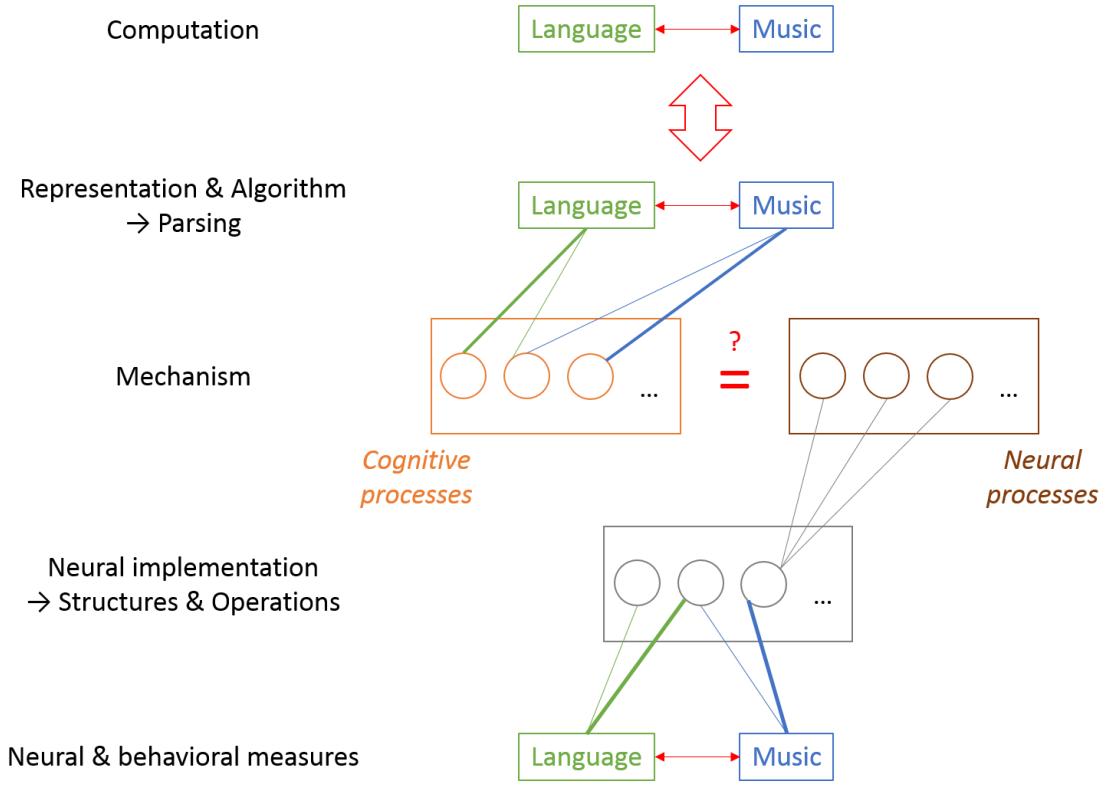


Figure 6.1 Levels of investigations in the comparative biomusicology framework.

As Figure 6.1 shows, there are top-down and bottom-up strategies to yield mechanistic explanations in terms of cognitive and neural processes. First, theory formulated explicitly at the computational and algorithmic level can contribute to improve $P(COG_X|TASK_Y$ and $STIMULUS_R)$ from top-down. Second, $P(ACT_Z|COG_X)$ should be optimized from bottom-up, i.e., from neural and behavioral measures to cognitive processes. As discussed above, the network approach to the brain and cross-disorder perspectives are two methods that optimize this side of inference (Poldrack, 2006; Price & Friston, 2005) which I also adopt in the current thesis. In addition, I suggest that implementational level research on neurobiological details reveal a set of neural processes which contribute to reveal cognitive processes (putatively) measured in experiments. In this way, I attempt to approach Type II explanation, i.e., explaining why certain NB objects are optimally designed to implement particular procedures (Embick & Poeppel, 2015), by integrating both top-down and bottom-up research strategies, and contribute to research on cognitive / functional ontology (Poldrack, 2006; Price & Friston, 2005) by identifying systematic relationship between cognitive and neural processes.

How is it possible to conduct comparative research within this framework? To achieve non-contrastive approaches, I put forward principled explanations regarding cognitive systems as instances of the same principles. This line of approach was also emphasized by several

researchers in the past (e.g., Merker, 2002; Merker, Morley, & Zuidema, 2015), but was not developed extensively in a comparative, biological context. One modern comparative, biological research program based on such an approach was introduced by Fitch (2014).[13] The central working hypothesis of the current thesis is as follows: Language and music share a set of basic computational and neural principles, but differ in their degree of expressions on the motor to cognitive gradient. The main goal of the current thesis is, then, to derive hypotheses about cognitive and neural processes of the music capacity and clarify their relation to those of the language capacity. At the best, those hypotheses are explicit enough to be implemented as a computational neurocognitive model and to generate implications which can be tested empirically in the future research. In the current thesis, I discuss how language and music can be explained in a principled way from two perspectives: 1) syntax as a set of combinatorial principles generating hierarchically structured representations and 2) neural structures and operations as implementational principles.

"Syntax" in its broad sense can be understood as "a set of principles governing the hierarchical combination of discrete structural elements into lager units (Bickerton, 2009) and/or into sequences (Patel, 2008)" (Asano & Boeckx, 2015, p. 2). Bickerton's way of the characterization is at the computational level, while Patel's way is relating to parsing. That is, the definition of "syntax" in its broad sense as introduced here subsumes both competence and performance aspects. This broad sense of "syntax" can be applied to language at different representational levels such as phonology, morphology, narrow-sense syntax, semantics, and pragmatics as well as to music in terms of tonal hierarchy and rhythmic structure. Hierarchical representations of language and music are different. However, once we take a bottom-up perspective and regard language and music as different instances of shared combinatorial principles, such differences can be explained in terms of difference in their goals to be achieved, i.e., the problems to be solved by the algorithm. As mentioned above, at the computational level, language is understood as mapping sound and meaning, while music as mapping sound and affect. However, both have combinatorial principles in common to achieve their specific goals. Therefore, syntax in the broad sense provides a good comparative basis.

Shepherd (2004) points out that investigations of organism's behavior and its biological substrates include multiple levels of organization: behavioral systems, interregional circuits, local (regional) circuits, neurons, dendritic trees, synaptic microcircuits, synapses, molecules and ions, and genes. As he suggested (Shepherd, 2004, p. 7), "[a]n important aim of the study

[13] I will come back to the approach introduced by Fitch (2014) in PART IV.

of synaptic organization is to identify the type of circuits and the functional operations that they perform at each of these organizational levels". At first glance, this approach might seem infeasible given the great varieties of neural organization and the vast amount of information about the brain. However, as Shepherd (2004) demonstrated, there are some basic principles common to the different brain regions and identifiable at successive levels of organization: there are fundamental types of synaptic circuits (canonical circuits) and elementary operations they perform (canonical operations). Those canonical circuits and operations serve as a conceptual framework for understanding how general principles of neural organization are adapted to achieve unique aspects of each brain region. Thus, neuroscientific endeavor investigating the neural basis of behaviors or cognitive systems should consider a set of principles underlying the varieties the brain organization seriously.

To make a step toward a computational neurocognitive model, in the current thesis, I derive a neurocognitive model from the results of comparative language-music research in terms of syntax as combinatorial principles and neural structures and operations as implementational principles. Several researchers suggested neurocognitive models of speech as well as language processing (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2002, 2011, 2012, 2016, 2017; Hagoort, 2013, 2016; Hickok, 2012; Hickok & Poeppel, 2007, 2016, 2004). In parallel, neurocognitive models of music processing were suggested (Koelsch, 2011a, 2012a; Patel, 2003, 2008, 2013). On one hand, such models constructed on the basis on language models could lead to a to-be-avoided contrastive approach. On the other hand, if language and music rely on the same neural structures and circuits as well as operations, neurocognitive models of language, which are already quite clear about the details of cognitive and neural processes, can contribute to build neurocognitive models of music processing. An example of such an enterprise is one suggested by Koelsch (2011a, 2012a) on the basis of Friederici's model. Thus, in the current thesis, I mainly focus on their models first and then discuss those models in light of other models to suggest some extension.

Friederici's model is a descriptive, boxological neurocognitive model for stepwise parsing of linguistic sentences, which was constructed on the basis of behavioral, brain imaging, and neuropsychological studies. However, it does not specify how the different sub-processes function (Jacobs & Hofmann, 2013). Therefore, the extended model which is introduced in the current thesis aims at providing mechanistic explanation in terms of cognitive and neural processes, which is explicit enough to be implemented as a formal-mathematical model in the future research. This direction is in line with the proposal that cognitive models should help us to specify hypotheses about the functions of brain activities for cognitive processes, so that

those hypotheses are falsifiable by the methods of cognitive neuroscience (Jacobs & Hofmann, 2013). Thus, though I do not build a formal-mathematical model in the current thesis, I will suggest a possible way toward computational neurocognitive modeling in cognitive musicology research.

PART II: Top-down perspectives to syntax and its neurocognitive mechanisms

The goal of this part is to investigate the relationship between language and music in terms of syntax as a set of combinatorial principles. In addition, shared aspects of syntax in language and music in terms of neurocognitive mechanisms, i.e. cognitive and neural processes, are clarified. I review the current comparative language-music research landscape concerning syntax from two perspectives: generative neurolinguistics and neuromusicology as well as neurocognitive psycholinguistics and psychomusicology (see Section 7.1, p. 37). I then introduce a first step toward tackling the problem of contrastive language-music research and the problem of explanatory gaps. The approaches of this part are characterized as top-down because they investigate syntax and its neurocognitive mechanisms on the basis of the concepts from ready-made computational-representational theories.

7 On comparing language and music

7.1 Syntax in the broad sense

As already mentioned in PART I, the term "syntax" is used in a broad sense (for details, see Asano & Boeckx, 2015) in the current thesis. In a broad sense, "syntax" can be defined as "a set of principles governing the hierarchical combination of discrete structural elements into lager units (Bickerton, 2009) and/or into sequences (Patel, 2008)" (Asano & Boeckx, 2015, p. 2). The broad sense of syntax applies to several cognitive domains such as language, speech, music as well as action, and opens up a new possibility for comparative research with a wide range of methods such as theoretical, psychological, and neuroscientific approaches. Below, I use "syntax" in this broad sense and call combination of words into phrases building up sentences "narrow-sense syntax". Syntax is one possibility (besides others such as Markov-chain-based or schema-based approaches) to deal with the problem of temporal integration and is applicable for structured sequential behaviors such as language and music as I will show in the current thesis. As repeatedly suggested, combinatorial property is not limited to narrow-sense syntax, i.e., combining words into sentences, but can be applied to a broad range of cognitive domains (e.g., Asano & Boeckx, 2015; Fitch, 2010a; Fitch & Martins, 2014; Fujita, 2016; Jackendoff, 2011, 2015).

In comparative research on language and music, there are a number of theoretical as well as empirical investigations based on such a broad understanding of "syntax" (Table 7.1). Table 7.1 contains theoretical and empirical approaches to language (left) and music (right) which are relevant for the discussion of the current thesis. The correspondence between research on language and music is indicated by locating them in the same row. The upper half of the table shows theoretical approaches, while the lower half depicts empirical ones. Theoretical and empirical research are not independent of each other. For example, cognitive neuroscience of language by Angela Friederici is based on the generative linguistic endeavor of Noam Chomsky. Neurocognitive model of music perception by Stefan Koelsch relates to the generative approach to music introduced by Martin Rohrmeier. Friederici and Koelsch take linguistic and musicological theories seriously and attempt to relate the theoretical concepts to neuroscientific research (i.e., direct correlational mapping strategy). Thus, I call those approaches "generative neurolinguistics and neuromusicology". The others prefer more unified theories of linguistics and psychology, and apply them to neuroscientific research (i.e., convergence strategy). Thus I call those approaches "neurocognitive psycholinguistics" and the counterpart in music by Aniruddh Patel "neurocognitive psychomusicology".

Table 7.1 Theoretical and empirical approaches to language and music

Linguistic theories	**Theories of music**
Generative grammar (e.g., Chomsky, 1956, 1957, 1965)	An unanswered question (Bernstein, 1976) & Generative Syntax of Tonal Harmony (Rohrmeier, 2007, 2011)
Minimalist Program (e.g., Chomsky, 1995)	Identity thesis (Katz & Pesetsky, 2011) and others (Asano, 2012; Fujita, 2016; Roberts, 2012)
Parallel architecture (e.g., Jackendoff, 2002)	A Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1983)
Psychology and neuroscience of language	**Psychology and neuroscience of music**
Brain basis of language processing (e.g., Friederici, 2002, 2011, 2017)	Neurocognitive model of music perception (Koelsch, 2011a)
Dependency locality theory (Gibson, 1998, 2000)	Shared syntactic integration resource hypothesis (Patel, 2003, 2008, 2012)
Memory, Unification, Control (e.g., Hagoort, 2005, 2013, 2016)	Shared syntactic integration resource hypothesis (Patel, 2003, 2008, 2012)

As I will review later, research on the relationship between syntax in language and music is very much influenced by difference in theoretical orientations mentioned above. However, the most important function of syntactic computation which seems common beyond theoretical borders is the existence of hierarchical structure underlying temporal sequence. In neuroscientific literature, Broca's region often tends to be associated with processing of hierarchical structures in several domains such as language (Friederici, 2011, 2012), music (Koelsch, 2011a, 2012a), and action (Fitch & Martins, 2014). Based on this general idea, a core function of syntactic computation can be characterized at the best as mapping between hierarchically structured representation and temporal sequence (see Figure 7.1). Figure 7.1 illustrates a function of syntactic computation in language / music as mapping between hierarchical structure and temporal sequence to link sound and meaning / affect. In comparative research on language and music, such a mapping should be investigated at the computational, algorithmic, and implementational levels (see Section 7.2, p. 39, Section 7.3, p. 41, and Section 7.4, p. 44, for more details).

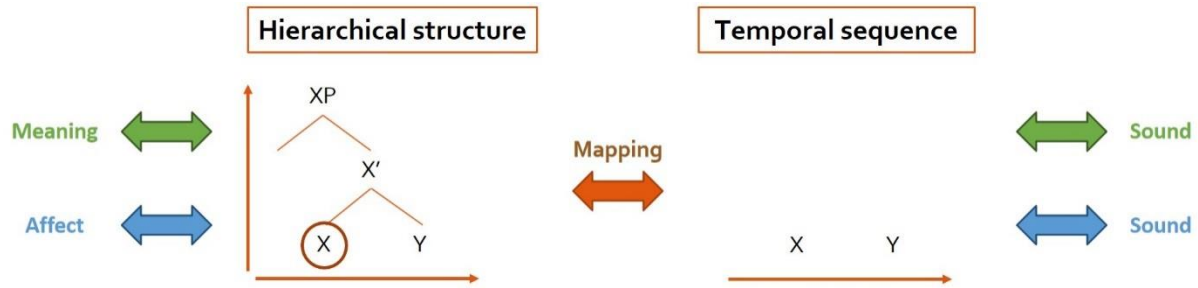


Figure 7.1 Syntax as mapping between hierarchical structure and temporal sequence

The main idea concerning the relationship between syntax in language and music, which I will introduce below, is that hierarchically structured representations of each cognitive system are different as the building blocks and computational goals of each cognitive system are different, but the fact that they are organized in headed hierarchical structures and must be mapped onto temporal sequences or vice versa is in common (see for example, Asano & Boeckx, 2015; Jackendoff, 2009; Jackendoff & Lerdahl, 2006; Patel, 2008; Thompson-Schill et al., 2013). Thus, comparative research on syntax in language and music should focus on such a mapping, regardless of whether the inquiry is at the computational level about the capacity / knowledge of language and music or at the algorithmic and implementational levels about moment-by-moment integration of events in time. In this way, language and music can be investigated as different instances of the same principles.

7.2 Computational level: Recursion and hierarchical structure building

At the computational level, language and music might differ from each other as their computational problems vary at first glance: linking sound and meaning for language and linking sound and affect for music. Hierarchical structures of the narrow-sense syntax in language are built on the basis of syntactic categories (e.g. Verb and Noun), lexical items, and propositional meaning. None of them has any proper parallel in music. Instead, hierarchical representations of musical syntax are non-categorical, encoding affect, and built on the basis of tonal and rhythmic stability or importance. Musical syntax also differs from phonological syntax. Hierarchical representations of music and phonology can be compared in terms of pitch and rhythm, but tonal hierarchy and isochronous meter are unique organizing principles of musical syntax. Therefore, language and music might differ in the way they satisfy those domain-specific conditions. However, at the same time, a computational problem for syntax in language and music is mapping between hierarchical structure and temporal sequence to link sound and meaning / affect. Thus, language and music might share some principles underlying

such a mapping. This hypothesis makes comparative language-music research at the computational level possible. Generative approaches to language and music (see Chapter 8, p. 46) which take recursion and hierarchical structure building as central aspects belong to this kind of comparative endeavor.

Recursion is a controversial and much debated term since Hauser and colleagues (2002) introduced it in the evolutionary research (see Fitch, 2010b; Lobina, 2017; M. D. Martins, 2012; Pinker & Jackendoff, 2005). However, as Lobina (2017) noted, it is the property of self-reference (or self-call) that subsumes all correct uses of recursion in the formal sciences. For example, the following definition of the factorial is recursive because "the recursive step involves another invocation of the factorial function […]" (Lobina, 2017, p. 11).[14] That is, the factorial function is invoked on both sides of the equation, i.e., the factorial function calls the factorial function itself.

$$(7\text{-}1) \quad Def.n! \begin{cases} if\ n = 1 & n! = 1 & (base\ case) \\ if\ n > 1 & n! = n \times (n-1)! & (recursive\ step) \end{cases}$$

In linguistics, as Bickerton (2009, p. 6) puts it, "recursion is generally defined as the ability to insert one structure inside another of the same kind". This characterization also takes the central property of recursion, i.e., self-reference, to heart. Recursion, in this way, can yield *discrete infinity* (Hauser et al., 2002, p. 1571). In other words, it is *theoretically* possible to define a countably (or denumerably) infinite set of discrete elements such as strings of symbols from a finite number of rules and a finite number of symbols. Especially, the view of language and music as Humboldt systems, i.e., infinite use of finite means, highlighted the necessity of theories accounting for discrete infinity (e.g., Chomsky, 1968, p. 15; Merker, 2002, p. 4; Merker et al., 2015, p. 3; Rohrmeier & Pearce, 2018, p. 476).

Hierarchical structure building is another central aspect of syntax in language and music. It was also claimed to be the most significant aspect of the narrow-sense syntax (Bickerton, 2009). For example, in language, relationship between words in a sentence, e.g., a subject-verb relationship, is determined not by linear order, but hierarchical structure. This can be illustrated by the following example.

[14] Given n = 4, according to the recursive step in the recursive definition n! = n x (n – 1)!. The first function application results in 4! = 4 x 3!, which means that the function has to evaluate 3!. This, in turn, is again the recursive step. Thus, the second function application results in 3! = 3 x 2!, so we get 4 x 3 x 2! and the third recursive step leads to 2! = 2 x 1!, i.e., 4 x 3 x 2 x 1!. Because then the base case applies with 1! = 1, we get 4 x 3 x 2 x 1, after evaluating this expression we get 24, i.e., the result of 4!.

(7-2) Everyone who knows J. S. Bach likes music.

In this sentence, “Bach” and “likes” are adjacent in linear order and appear to be in the range of the subject-verb agreement. However, in the hierarchical structure indicated by brackets, those two elements are ‘far away’.[15]

(7-3) [[[Everyone] [who [knows [J. S. Bach]]]] [[likes] [music]]].

Similarly, relationship between pitch events in a tonal sequence, e.g., tonic-dominant relationship, is determined not by physical distance, but cognitive distance based on the tonal hierarchy (Jackendoff & Lerdahl, 2006). Therefore, in both language and music, theories of hierarchical structure building are required to explain the cognitive relationship between elements. Moreover, hierarchical structure of language and music not only stands for hierarchical relationship between elements (called *dominance relationship*), but also includes a *head* element, the red circle in the Figure 7.1, which determines the *label* of each hierarchical unit.

Generative approaches[16] to language and music then investigate hierarchical structure building by taking domain-specific constraints into account. By means of recursion and hierarchical structure building, most of those approaches[17] attempts to characterize all theoretically possible grammatical strings in language and music, and explain structural relationships between elements. In addition, the relationship between surface configurations and argument structure in language or tension-relaxation pattern in music is determined by a set of rules, constraints, or templates, which are also objects of research on syntax. Examples of theories applying recursion and hierarchical structure building for language and music are discussed in the Chapter 8 (p. 46).

7.3 Algorithmic level: Parsing models

At the algorithmic level, mapping between hierarchical structure and temporal sequence is often investigated as a core function of parsing, i.e., structure analysis during moment-by-moment integration of events in time. Again, algorithms in language and music

[15] The distance in the hierarchy can be intuitively followed in terms of the number of the brackets between “Bach” and “likes”.
[16] Generative approaches to language and music as used in the current thesis include not only computational level investigations, but also representational theories (i.e., one part of algorithmic level investigations).
[17] *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983) is an exception as it rather deals with a psychologically-based parsing theory of temporal organization. See also Footnote 22.

may use different representational formats and thus differ to some degree. For example, linguistic parsers should deal with phonological, morphosyntactic and phrasal syntactic, semantic as well as pragmatic information. A musical parser should process tonal-harmonic and beat-based rhythmic information to encode affect. However, the mapping between hierarchical structure and temporal sequence remains the central problem for a linguistic and musical parser to solve. Moreover, as indicated in the Figure 7.1, hierarchically structured representations of language and music determine hierarchical relationship between elements (i.e., dominance relationship) and the head (i.e., an element identifying the label of each hierarchical unit). Both language and music also contain a temporal dimension. That is, in both cases, the algorithm should deal with headed hierarchical structure and temporal sequence.

Psychology and neuroscience of language and music are often concerned with real-time processing of sequences in terms of parsing. A parser is an algorithm realizing online structure analysis. In investigating parsing, it is useful to distinguish different parsing models (Harley, 2001). First, the distinction should be made between autonomous and interactive models. While autonomous models assume the initial modular stages of parsing which rely solely on syntactic information, interactive models undertake constant influence by multiple information sources (e.g., semantic information) on the syntactic processor. Second distinction is concerned with one-stage and two-stage models. While one-stage models use syntactic and semantic information in one pass, two-stage models consist of an autonomous stage of syntactic processing followed by the second stage where semantic information is also used. Moreover, fixed- and variable-choice models should be differentiated. The fixed-choice models are also characterized as deterministic because they don't make use of probabilistic elements in choosing structural analysis. As they choose one structural analysis at an early time point, they require reanalysis if the structural analysis turns out to be incorrect at a later stage. On the contrary, the variable-choice models decide on a final analysis on the basis of a competition between alternative analyses, i.e., a winner-take-all principle.

There are mainly three syntactic phenomena which comparative research in psychology of language and music frequently investigates in light of parsing: grammaticality or acceptability, complexity, and ambiguity. Those phenomena are explained differently according to which type of parsing model is assumed (for reviews, see Harley, 2001; Jackendoff, 1991; Sprouse & Lau, 2013). For example, an autonomous two-stage fixed-choice parser assumes that syntactic information of the sentence is first processed alone, which determines

the structural analysis on the basis of principles such as minimal attachment and late closure[18] before semantic information enters into the analysis. An incorrect analysis then leads to a reanalysis. This type of parsing models is called 'syntax-first'. In contrast, an interactive one-stage variable-choice parser uses multiple sources of information called constraints in parallel, keep alternative structural analyses, and operates on the basis of the principle of competition. Those two parsing models are at the different end of a theoretical pole. In addition, there are also parsing models which make use of psychological concepts such as working memory and cognitive control.

As Harley (2001, p. 263) puts it, "[a]ny account of parsing must be able to specify why sentences are assigned the structure that they are, why we are biased to parse structurally ambiguous sentences in a certain way, and why some sentences are harder to parse than others". All above mentioned parsing models attempt to satisfy this requirement and it is very hard to decide between alternative models. Selection of a parsing model is often influenced by assumptions from a particular computational-representational theory. On one hand, the tight link between computational-representational theory and parser, i.e., algorithm, seems striking because investigations at the algorithmic level tend to start with an assumption about representations which must be built and units which are manipulated. On the other hand, however, as Marr (1982) also noted, the same computational problem can be solved by different algorithms, indicating no direct mapping between computational-representational theory and parser.

In sum, there are several ways to approach mapping between hierarchical structure and temporal sequence in terms of parsing. While the goals of parsing in language and music are partially different, the parser in both case has to deal with the problems such as grammaticality, complexity, and ambiguity in mapping between hierarchical structure and temporal sequence. A 'syntax-first' model of parsing in language and its relation to musical parsing is discussed in Chapter 9 (p. 56) from the perspectives of generative neurolinguistics and generative neuromusicology, which is then discussed in relation to further models in Chapter 10 (p. 85).

[18] Minimal attachment and late closure are parsing principles introduced by Frazier and Fodor (1978). It is suggested that they are utilized to quickly and efficiently assign structure to incoming sentences. On the basis of minimal attachment a structural analysis which contains the fewest number of nodes should be adopted. Late closure states that incoming elements should be integrated into the phrase or clause which is currently processed.

7.4 Implementational level: Fronto-temporal networks

The implementational level, then, is concerned with how and where parsing is realized in the brain. As I will discuss in Chapter 9 (p. 56) and Chapter 10 (p. 85), research at the implementational level depends on the parsing model used to investigate the mapping between hierarchical structure and temporal sequence. Depending on which model one chooses, interpretation of the results in cognitive neuroscience differs (Kaan, 2007; Sprouse & Lau, 2013). In this case, a particular parsing model is mapped onto the brain. For example, such a mapping is investigated in terms of hierarchical processing and executive function. Nevertheless, there is a set of neural structures suggested to be relevant to syntax beyond differences in parsing models, namely the left dorsal fronto-temporal network including subparts of the Broca's region and the posterior superior temporal lobe as well as the arcuate fasciculus (AF). Before I will discuss those structures in Chapter 9 (p. 56) in details by referring to findings from generative neurolinguistics and neuromusicology, below I sketch some aspects of those structures which yielded some consensus to take them as starting points of discussions. I particularly focus on the Broca's region here as it was suggested repeatedly as a core shared structure for syntax in language and music (Fitch & Martins, 2014; Koelsch, 2011a, 2012b, 2012a; Patel, 2008, 2012).

Broca's region is located in the left inferior frontal gyrus (IFG). The relatively simple cytoarchitectonic subdivisions of the Broca's region in the Brodmann area (BA) 44 and 45 are well-known (Figure 7.2a). The BA 44 and 45 are not only cytoarchitectonically different, but also differently connected with other regions of the brain via long fiber tracts, i.e., bundles of axons (for a review, see Petrides, 2014, pp. 151–154). One of the long fiber bundles well-known in language research is the arcuate fasciculus (AF) which monosynaptically connects Broca's region to the posterior superior temporal region. The AF consists of three branches which were found in macaques and interpreted as homologous connections in the human brain as shown in the Figure 7.2b (for discussions, see Petrides, 2014). Figure 7.2b also shows a complex relation between the AF and the middle longitudinal fasciculus (MLF). The fibers of the AF arch around the end of the lateral fissure and extend to the region where the MLF runs. The complexity even increases once the superior longitudinal fasciculus (SLF) I, II, and III are taken into account as they overlap largely with the AF (see for example, Gierhan, 2013; and Petrides, 2014, p. 171), but I do not elaborate this issue in the current thesis.

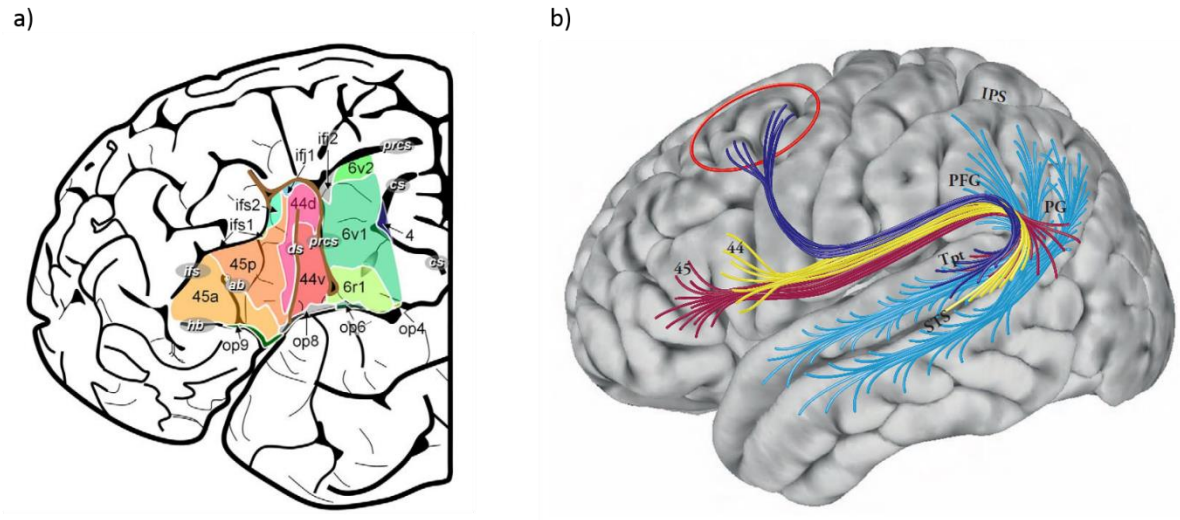


Figure 7.2 a) Broca's region (BA 44 and 45), premotor cortex (BA 6), and other surrounding structures adopted from Amunts (2010), *PLoS Biology*, 8: e1000489; b) Arcuate fasciculus (red, yellow, and dark blue) and middle longitudinal fasciculus (light blue) adopted from Petrides (2014), *Elsevier Books*, *Neuroanatomy of Language Regions of the Human Brain*, with permission from Elsevier.

As I will review in the Chapter 9 (p. 56), a branch of the AF originating from the left posterior superior temporal gyrus / sulcus and terminating in the left BA 44 is ascribed to a pathway critical to syntax in language. Does this branch of the left AF also play a crucial role in music? In addition, concerning Broca's region, as shown in Figure 7.2, a further subdivision was introduced on the basis of the difference in the regional distribution of transmitter receptors (Amunts et al., 2010). The BA 44 is segmented into a dorsal and a ventral part, while the BA 45 is partitioned into an anterior and a posterior portion. Amunts and colleagues (2010) note that the areas involved in similar functions have similar receptor distribution patterns. What is the consequence of this new parcellation regarding the claim about Broca's region as a core shared structure for syntax in language and music?

Experiments which provided evidence for the hypotheses about shared neural resources for syntactic processing in language and music (e.g., Koelsch, Gunter, Wittfoth, & Sammler, 2005; Kunert, Willems, Casasanto, Patel, & Hagoort, 2015; Sammler et al., 2009; Slevc, Rosenberg, & Patel, 2009; Steinbeis & Koelsch, 2008) regard interaction and neural overlap as implications for neural sharing on the basis of following inferences: 1) Sharing → Interaction; 2) Sharing → Overlap. Especially, I will review and discuss neural overlap as an implication of sharing by focusing on the fronto-temporal networks and especially on the Broca's region later.

8 Generative approaches to language and music

8.1 Formal language theory and the Minimalist Program in generative linguistics

One approach investigating syntax is about theories of generating and building hierarchical representations underlying sequences, with its primary goal as explaining the mystery of language acquisition: children are able to acquire knowledge of language in its full richness, complexity, and specificity in spite of the limitation of the data available (Chomsky, 1986, p. 55). Thus, the target of inquiry is the set of principles generating the knowledge or internal representations of cognitive systems. In grammar approaches, the knowledge is represented by formal rule systems (i.e. grammars) generating all strings of symbols (i.e. sentences or expressions) which belong to a formal language (for details about the formal language theory, see, e.g., Jäger & Rogers, 2012; Partee, Ter Meulen, & Wall, 1993). In this context, a grammar is understood as a string rewriting system, consisting of the start symbol (usually S) and a finite number of rules in form of $\psi \rightarrow \omega$, which means that a string (ψ) is replaced by another string (ω) to yield a new string. Grammars use two kinds of alphabets: a terminal alphabet (usually small letters such as "a" and "b") and a non-terminal alphabet (usually capital letters such as "A" and "B").

In the same way, a natural language can be also regarded as a set of strings (e.g., sentences). The way of applying the rules is reflected in the derivation of the strings, which can often be represented by a constituent structure tree, i.e. a hierarchically structured representation of a string (Figure 8.1).

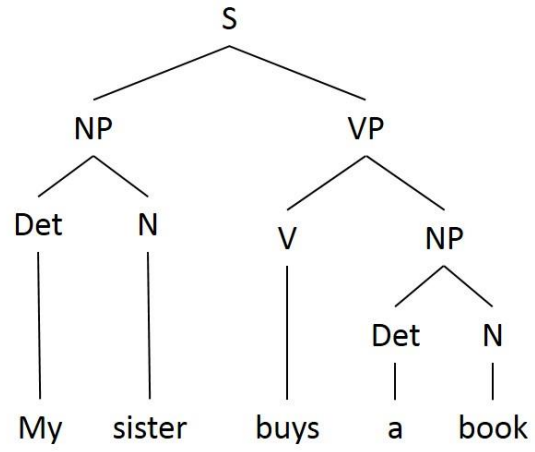


Figure 8.1 Constituent-structure tree of a sentence.

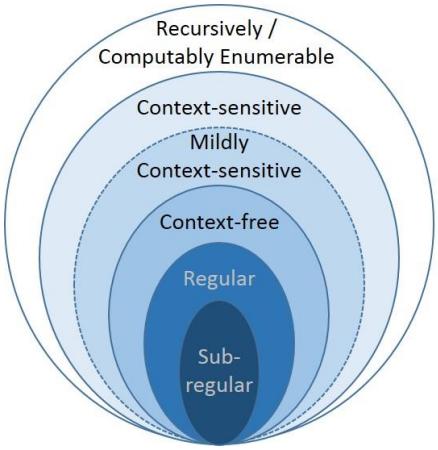
This tree represents three sorts of information about the syntactic structure of a sentence: 1) hierarchical organization of the units into constituents (i.e., the dominance relation

of constituents), 2) the syntactic category of each constituent such as N (noun) and V (verb) by labeling each node, and 3) the serial order of the constituents (i.e. the precedence relation of constituents) (Partee et al., 1993, Chapter 16). For example, in Figure 8.1, the node labeled VP (verbal phrase) dominates the lower nodes such as V and NP (nominal phrase) as well as Det (determiner) connected to “a” and N connected to “book”, but does not dominate the other NP consisting of Det connected to “My” and N connected to “sister”. Two nodes (e.g. V and NP) immediately dominated by the same node (e.g. VP) are called sister nodes and ordered in the left-to-right fashion: the node V precedes NP. Such a linguistic constituent structure tree is also called “phrase-structure tree”. The tree in Figure 8.1 is generated by the following phrase structure grammar:

(8-1) S -> NP VP
(8-2) NP -> Det N
(8-3) VP -> V NP
(8-4) Det -> my | a
(8-5) N -> sister | book
(8-6) V -> buys

According to such a grammar, it is possible to decide whether strings belong to a particular language and thus ‘grammatical’ or not.

Importantly, there are several formal grammar types, which differ from each other in terms of their generative capacity. In parallel, there are different types of formal languages. The relationship between different grammar and language types is given by the Chomsky hierarchy which includes four classes: regular (or type-3) grammar and language, context-free (or type-2) grammar and language, context-sensitive (or type-1) grammar and language, and type-0 grammar and recursively (or computably) enumerable language (Figure 8.2).



Language	Grammar
Recursively enumerable Computably enumerable	Type-0
Context-sensitive	Context-sensitive Type-1
Context-free	Context-free Type-2
Regular	Regular Type-3

Figure 8.2 (Extended) Chomsky hierarchy.

The type-3 grammar includes the most constrained form of the rules and thus is most limited in its generative capacity, while the type-0 grammar does not include any restrictions on the form of the rules and thus is most powerful in its generative capacity. The restrictions of rules can be stated as following:[19]

(8-7) Type-0: $\psi \rightarrow \omega$.

(8-8) Type-1: $\alpha \, \mathrm{A} \beta \rightarrow \alpha \, \psi \, \beta$, where $\psi \neq \mathrm{e}$.

(8-9) Type-2: $\mathrm{A} \rightarrow \psi$.

(8-10) Type-3: $\mathrm{A} \rightarrow x \, \mathrm{B}$ or $\mathrm{A} \rightarrow x$.

In all types of grammar, the left side of the arrow cannot be empty and contains at least a non-terminal symbol. Type-0 grammar can be understood as a string rewriting system without any other constraints. Type-1 grammar allows for rewriting a string only in a certain context ($\alpha \_ \beta$), where the string on the right side of the arrow cannot be shorter than that on the left side. This also means that the Type-1 grammar does not permit to replace a string through an empty string (e). In the Type-2 grammar, the left side of the arrow should be a single non-terminal symbol, which can be replaced by any string, i.e., a string mixing up terminal and non-terminal alphabets. Type-3 grammar, in addition, restricts the right side of the arrow such that a string x of terminals, is followed by at most one non-terminal symbol. Grammars modeling the structural complexity of natural language sentences are called "mildly context-sensitive grammars" (Jäger & Rogers, 2012). This class forms a subset of the class of Type-1 grammars.

[19] The following descriptions and explanations of different types of grammar are summarized on the basis of Partee et al. (1993, p. 451).

The type-2 grammar, i.e., context-free grammar, plays a significant role in generative neurolinguistics as well as comparative cognition research. Thus, it earns more detailed illustration here. The context-free grammar gained attention especially since Chomsky's finding that English is not a regular language because of its center embedding construction where the embedding is potentially unbounded (Chomsky, 1957). To characterize such a language, at least, a context-free grammar is required. For example, an English embedded relative clause sentence (e.g., "The mouse the cat the dog loves married run away."[20]) can be created by the grammar with the rules S -> NP S VP and S -> NP VP which allow potentially unbounded embedding of a relative clause into another relative clause. The language characterized by this kind of grammar is also called A^nB^n-language.

Although the grammar approach is often applied in linguistic research to generate sentences (i.e., strings of words) belonging to a language, natural language contains further strings such as sequences of morphemes or phonemes (Partee et al., 1993). Thus, phrasal, morphological, and phonological syntax all fall into the scope of the grammar approach. To generate strings in different domains of language, grammars with different generative capacity are suggested to be necessary: the generative capacity of at least the type-2 grammar (and type-1 grammar for some cases) is required for sentential syntax, at least the type-3 (and type-2 grammar for some cases) for morphology, and at most type-3 (and maybe largely limited to its subset called sub-regular grammar) for phonology (Heinz & Idsardi, 2013).

The Minimalist Program (Chomsky, 1995), i.e., a modern approach within the generative linguistic tradition, is arranged at the level of mildly context-sensitive grammar and focuses on a minimal condition for linguistic discrete infinity called "Merge". Merge takes two already formed structures like X and Y, and combines them hierarchically into a new structure A = {X, Y} (Chomsky, 2010, p. 52). Further, two different types of Merge are differentiated: Given A, we can merge an element to it from outside of A (external MERGE) or from within A (internal MERGE; displacement). That is, external MERGE creates a structure like {B, {X, Y}} and internal MERGE creates a structure like {X, {X, Y}}. Therefore, Merge generates hierarchically structured representations, i.e., dominance relationship, but does not determine the label of the set nor the order of the elements, i.e., precedence relationship. In this approach, most of the complexity attributed to grammars in previous research is regarded as a by-product of interface systems such as the sensory-motor interface system (e.g., phonology) and the

[20] This sentence in the full length is "The mouse which the cat which the dog loves married run away." and is organized according the rules hierarchically as "$_S$[$_{NP}$[The mouse] $_{S'}$[which $_S$[$_{NP}$[the cat] $_{S'}$[which $_S$[$_{NP}$[the dog] $_{VP}$[loves]]] $_{VP}$[married]]] $_{VP}$[run away].]".

conceptual-intentional interface system (e.g., semantics and pragmatics) (Chomsky, 2010). The labeling is carried out by the general principle of minimal search, which determines a structurally more prominent element as a label (Fujita, 2016). Moreover, the mapping between hierarchical representations and linear order is executed by multi-step linearization operations (Boeckx, Martinez-Alvarez, & Leivada, 2014).

For the discussions below, further concepts such as head-complement as well as dependency and agreement are relevant. The first concept was originally introduced as an additional characterization / constraint for context-free phrase structure grammars: A phrase consists of a head with the same category type as the phrase and a complement.[21] Thus, concerning the example in the Figure 8.1 above, "book" with the category N is the head of the phrase NP, while "a" is the complement. The dependency, i.e., correlation between elements of a sequence, is one computational problem for syntax and the agreement is one example of a syntactic dependency (Kaplan, 2003, p. 74). Concerning the example in the Figure 8.1 again, "My sister" and "buys" are in a dependency relationship and there is an agreement between those elements in terms of grammatical number, i.e., singular.

8.2 Generative theories in musicology

8.2.1 Generative Theory of Tonal Music

The view that mental representations of music can be also investigated by grammars was held by several researchers especially in the 1980's (for reviews, see Rohrmeier & Pearce, 2018; Seifert, 1993; Sundberg & Lindblom, 1991) and culminated in A Generative Theory of Tonal Music (GTTM) by Lerdahl and Jackendoff (1983). Such a grammar approach was suggested as one of promising areas in cognitive musicology research, together with an artificial intelligence approach (Seifert, 1993). Although GTTM is situated in the generative grammar framework, it aimed at a psychological theory of temporal organization rather than a formal description of music.[22] Especially, their goal was to identify the knowledge of music in form of mental representations which are constructed by (experienced) listeners. Moreover, it assumes multiple generative systems working in parallel, as also seen in Jackendoff's constrained-based parallel architecture which regards syntax, phonology, and semantics as independent, but interactive generative systems (e.g., Jackendoff, 2002, pp. 107–151).

[21] This description is based on that made by Bußmann (2008, pp. 376 & 804) in her linguistic lexicon.

[22] Although the GTTM as a psychological theory of temporal organization which takes the convergence strategy (described in PART I, p. 23) rather corresponds to Jackendoff's parallel architecture, and thus differs from the remainder of the generative approaches, it is a reference theory for other theories investigating musical syntax and is thus reviewed as the first theory in the current section.

GTTM mainly consists of four sub-systems organizing pitch-events into hierarchically structured representations (Lerdahl & Jackendoff, 1983): meter, grouping, time-span reduction, and prolongational reduction. Metrical structure is the hierarchical organization of periodically recurring strong and weak beats. Grouping structure is the hierarchical organization of the musical stream into motives, phrases, and sections. Further, time-span and prolongational structures were introduced to account for the property of musical knowledge that some pitch-events are heard as ornamentations or elaborations of others. That is, pitch-events are heard in a hierarchy of relative importance. In the time-span structure, harmonic and melodic elaborations are studied in the rhythmic framework constructed on the basis of metrical and grouping structures. The prolongational structure, then, encodes the tension and relaxation patterns (called affect) across phrases, i.e., structural relationships which are non-adjacent, on the basis of melodic and harmonic stability conditions.

Those four components were extended on the basis of Lerdahl's tonal pitch space (TPS) theory (Lerdahl, 2001; for a short summary of the development of the GTTM and TPS, see Lerdahl, 2009). Especially, the details of the stability conditions were worked out by including components such as tonic finding and pitch-space distances (Figure 8.3). The basic idea of the TPS is that "the cognitive distance of an event from a given reference point measures the instability of that event in relationship to the reference point" (Lerdahl, 2009, p. 191). That is, once the reference point (e.g., the tonic or tonal center which is the most stable pitch-event in a Western tonal system) is identified, the stability (or rather instability) of the other pitch-events can be determined on the basis of the distance in a pitch space or tonal hierarchy. For example, Figure 8.3 demonstrates the basic space for the major and minor modes in Western tonal music (Lerdahl, 2001). Lerdahl (2001) conveys further pitch-class and chordal spaces from this basic space according to the rules stating how much pitch-class representation in the basic space should be shifted to attain a new space.

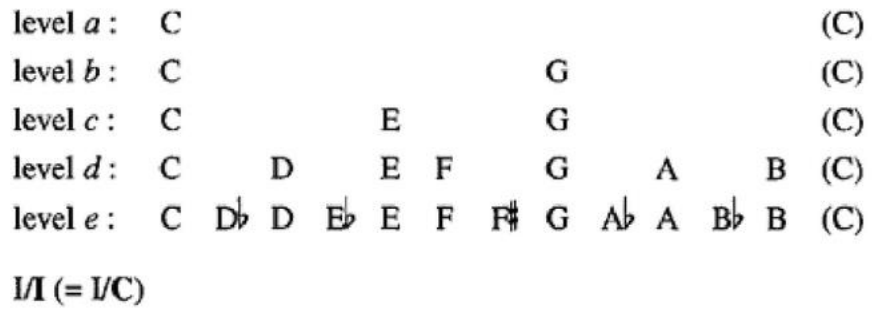


Figure 8.3 This figure shows the basic space with capital letters representing pitch-classes. In this figure, the tonal center is the key of C and pitch class C is on scale step I. The more stable

a pitch-class, the higher the level it is projected on to. This figure is adopted from Lerdahl (2001, p. 47) with permission by Fred Lerdahl.

Coming back to the GTTM, two central rule systems should be mentioned: "well-formedness rules, which specify the possible structural descriptions, and preference rules, which choose from the possible structural descriptions those that correspond to experienced listeners' hearings of any particular piece" (Lerdahl & Jackendoff, 1983, p. 9). Thus, well-formedness rules generated multiple, competing representations, of which one[23] representation is chosen as the most stable interpretation on the basis of preference rules. Lerdahl (2009) characterizes preference rules as a conceptual midpoint between neural networks, in which "a structure arises from the strongest activation in a network without reliance on rules per se" (p. 189), and optimality theory, in which "rule derivations are ordered and ranked according to a winner-take-all principle" (p. 189). The introduction of preference rules, therefore, made the GTTM significantly differ from the grammar approach in generative linguistics. Rather, preference rules are similar to one of heuristic strategies used in parsing theory, namely making a 'best guess' about the sentence structure (Lerdahl & Jackendoff, 1983, p. 314).

Another important differentiation made in the discussion of hierarchical structure in music is that of event hierarchy and tonal hierarchy (Bharucha, 1984; Deutsch, 1984; Dowling, 1984; Lerdahl, 2001; Patel, 2008; Tillmann, 2012). The former is a hierarchy of the string of musical events that constitutes a particular piece of music, while the latter hierarchy is not specific to each musical piece, but is a hierarchy of event classes encoding implicit knowledge of the abstract musical structure of a culture or genre (Bharucha, 1984). "A native listener has access to tonal hierarchies in long-term memory, which, once activated, can facilitate the generation of event hierarchies" (Bharucha, 1984, p. 425). That is, an event hierarchy integrates both tonal and temporal information to represent hierarchical relationship between events in time. Time-span and prolongational structures of the GTTM are examples of event hierarchies. A tonal hierarchy rather relates to stability conditions of the GTTM which represents long-term knowledge of a particular tonal system independent of the instantiation in a particular piece. The group of event classes that are stable within a given tonal hierarchy is also called "tonal schema" (Bharucha, 1984).

[23] The preference rules do not determine only one preferable representation unambiguously. However, according to the preference rules, representations are preferable in different degree so that some representation is more likely to be chosen than the others. Here, I say *one* representation is chosen as the most stable interpretation because several authors pointed out that only one representation is attended at a time (e.g., Jackendoff, 1991, p. 214; Temperley, 2000, p. 81; Vuust & Witek, 2014, p. 8).

8.2.2 Generative Syntax of Tonal Harmony

The most recent theory of musical grammar is Generative Syntax of Tonal Harmony (GSTH) by Rohrmeier (2007, 2011), which provides a formalization of hierarchical tonal-harmonic structure of music based on the Riemannian functional tradition in the way similar to the linguistic phrase structure grammar mentioned above. Rohrmeier's approach is based on formal grammar. In contrast to the GTTM, Rohrmeier's theory recursively enumerates a set of strings which belong to a particular set, i.e., a set of tonal harmonic sequences. Thus, according to this theory, there is such classification as grammatical and ungrammatical string sets.

The core idea of GSTH is that a harmonic sequence can be derived from three elementary harmonic functions, i.e. *tonic*, *dominant*, and *subdominant*, and thus can be regarded as fundamentally grounded in elaboration of cadential harmony. The formalism of this theory described in form of a context-free grammar distinguishes four levels: a phrase level, a functional level, a scale degree level, and the surface level. First of all, a piece in a certain key ($\mathbb{K}$ is a set of musical keys) consists of one or more phrases (P^+) such that each features a tonic region (TR) as its functional region. On this phrase level, the key of the piece is determined by means of this TR.

(8-11) $piece_{key=x \in \mathbb{K}} \rightarrow P^+$

(8-12) $P \rightarrow TR$

On the functional level ($\mathbb{R}$ is a set of functional regions including tonic region (TR), dominant region (DR), and subdominant region (SR)), three sets of rules are introduced. The first set of rules called "functional expansion rules" generates core functional sequences terminating with tonic (t), dominant (d) and subdominant (s).

(8-13) $TR \rightarrow DR\ t$

(8-14) $DR \rightarrow SR\ d$

(8-15) $TR \rightarrow TR\ DR$

(8-16) $XR \rightarrow XR\ XR$ for any $XR \in \mathbb{R}$

(8-17) $TR \rightarrow t$

(8-18) $DR \rightarrow d$

(8-19) $SR \rightarrow s$

The second set of rules called "substitution rules" allows for replacing or substituting each functional symbol by their relatives or parallels.

(8-20) $t \rightarrow tp$

(8-21) $t \rightarrow tcp$

(8-22) $s \rightarrow sp$

(8-23) $d \rightarrow dp$

The last set of rules called "modulation rules" enables each function ($\mathbb{F}$ is a set of functional terms including t for tonic, d for dominant, s for subdominant, tp for tonic parallel, sp for subdominant parallel, dp for dominant parallel, and tcp for tonic counter parallel) to become a local tonic.

(8-24) $X_{key=y} \rightarrow TR_{key=\psi(X,y)}$ for any $X \in \mathbb{F}$ and $y \in \mathbb{K}$

(8-25) $X_{key=y\ maj/min} \rightarrow X_{key=y\ min/maj}$ for any $X \in \mathbb{F}$ and $y \in \mathbb{K}$

On the scale degree level, the functional terms are replaced by scale degrees and relative scale degrees. On the surface level, the scale degrees are transformed into chords in terms of the given key property of the piece.

Figure 8.4 is an example musical constituent structure which is constructed solely on the basis of harmonic rules introduced above. First, this piece in G major key consist of two phrases according to rule (8-11). Those phrases correspond to tonic region (*TR*) according to rule (8-12). In focusing on the left hand tree, *TR* consists of *TR* and dominant region (*DR*) according to rule (8-15). The second *TR* then is composed of two *TR*s according to rule (8-16). The left hand *TR* again consists of two *TR*s which then terminate in tonic (*t*). The chord function is replaced by scale degree (roman number I for tonic) which turns into chord G major.[24] The right hand *TR* is composed of dominant region (*DR*) and tonic (*t*) according to rule (8-13). The *DR* is decomposed into subdominant region (*SR*) and dominant (*d*) according to rule (8-14). *SR* is then replaced by subdominant (*s*) according to rule (8-19). The *DR* in the middle represents application of modulation rules: *d* becomes a local tonic according to (8-24). Thus, *t* of this *TR* is D major. The rest follows the same procedure as I already explained.

[24] Below, I skip this step of explanation. In general, tonic (*t*) corresponds to roman number *I* which is G major, dominant (*d*) to *V* which is D major, subdominant (*s*) to roman number *IV* which is C major, and tonic parallel (*tp*) to *VI* which is E minor.

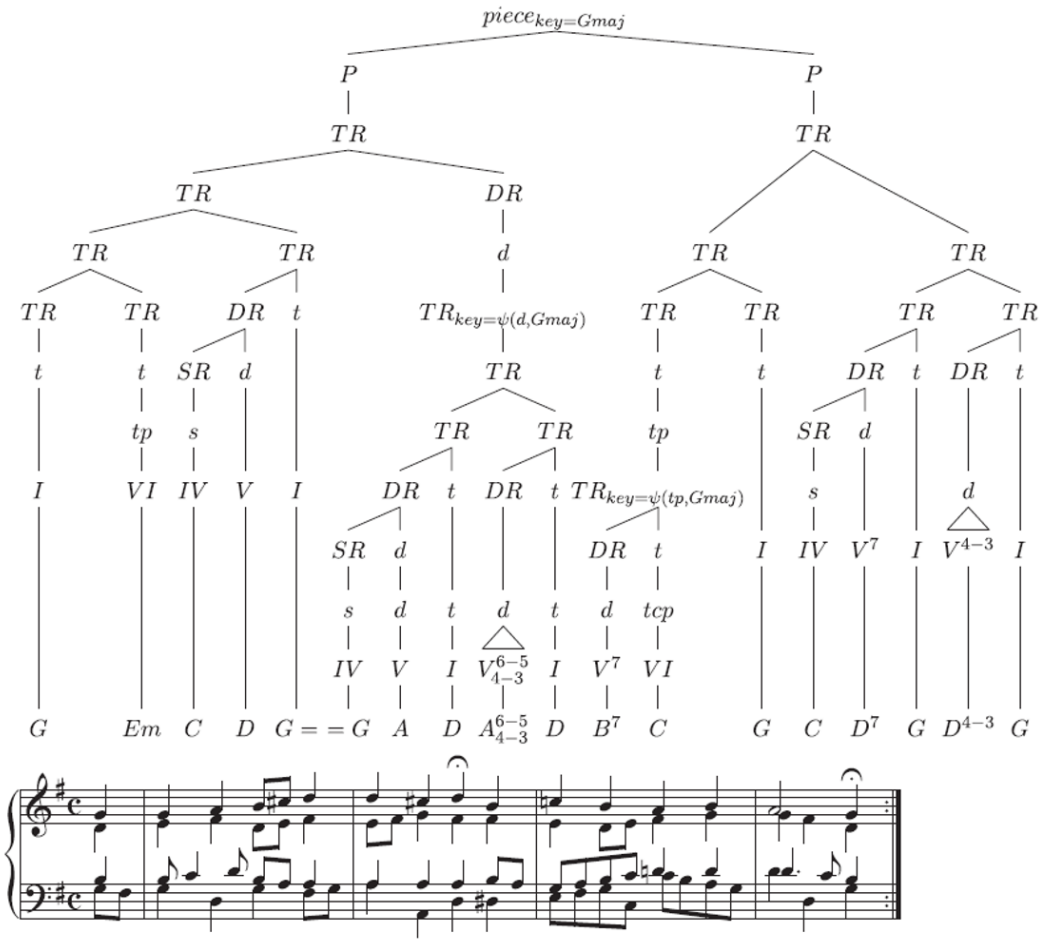


Figure 8.4 Musical constituent structure constructed on the basis of harmonic rules. This figure is adopted from Rohrmeier (2011), *Journal of Mathematics and Music*, 5(1): 35-53, with permission by Taylor & Francis.

There are two concepts central to Rohrmeier's theory. First, "each element (chord) in a chord sequence is structurally connected to its preceding or succeeding chord or chord group in a dependency relationship" (Rohrmeier, 2011, p. 38). Among two dependent elements, one is called "head" and the other is called "dependent". Importantly, the dependency relationship between two elements is established not only between adjacent elements (adjacent dependency), but also distant elements (long-distance dependency). Second, "chords are organized into functional categories which describe their tonal function which may be instantiated or modified by different chords" (Rohrmeier, 2011, p. 38). The functional categories include the three main tonal functions *tonic*, *dominant*, and *subdominant*. This second aspect plays a role in modulation. Modulation shown in the Figure 8.4 (indicated with "$TR_{key=\psi(d,Gmaj)}$") can be characterized as embedding of a key (e.g., D major) into another key (e.g., G major). In this case, there are two instantiations of the category *tonic*: one is D major and the other is G major.

In addition, in the Figure 8.4, D gets two categories: one is *tonic* in the local context of G major, the other is *dominant* in the global context of the G major. That is, the category of each chord depends on the current tonal context.

9 Generative neurolinguistics and generative neuromusicology

9.1 Angela Friederici's cortical circuit model of language

The most popular brain model of language processing in the generative linguistics tradition is Angela Friederici's cortical circuit model of language which explains on-line auditory sentence processing ranging from auditory perception to sentence comprehension in terms of a dynamic temporo-frontal network (Friederici, 2002, 2011, 2012, 2015, 2017; Friederici & Gierhan, 2013). This model is an autonomous model assuming initial modular stages of parsing which rely solely on syntactic information and assumes structure building, agreement checking, and thematic-role assignment as operations involved in syntactic computation. Moreover, it tells apart simple and complex syntactic processing. The particularity of the Friederici's model is mainly that it integrates time course and neuroanatomical basis of auditory sentence processing on the basis of neuroimaging experiments using different methods, and thus provides a hypothesis about when and where a particular aspect of language is processed in the brain. Major methodological paradigms used in those neuroimaging experiments comprise 1) violation processing, 2) artificial grammar learning, 3) isolating syntactic processing, and 4) manipulation of syntactic complexity. The first two methods relate to experimental paradigm, while the latter two are concerned with manipulation of the stimuli according to particular parameters. In this section, first of all, those methodological approaches are summarized with emphasis on syntactic processing, i.e., morphosyntactic and phrasal syntactic processing, as it is the focus of later discussion. Then, time course and neuroanatomical issues of the model are recapped.

9.1.1 Methodological paradigms

9.1.1.1 Violation paradigm

One of the most common approaches to test syntactic processing is the violation paradigm in which correct and incorrect natural sentences are contrasted. In most cases, incorrect sentences contain at least one of the syntactic violation types listed in the Table 9.1. The violation approach assumes that a sentence parser is equipped with a set of rules, by means of which irregularity can be detected. In the course of processing sentences in natural languages, a parser has to check different types of incoming information (Friederici & Weissenborn, 2007).

For example, word category information (e.g., verb, noun, determiner, and preposition) should be checked to build a local phrase structure (e.g., verb phrase, determiner phrase, and prepositional phrase). This relates to structure building operation. In addition, morphosyntactic information (e.g., inflection) and its agreement (e.g., subject-verb and determiner-noun agreement) as well as verb's argument structure (transitive or intransitive) should be identified and checked. This recruits agreement checking operation.

Table 9.1 Violation types and example sentences / phrases. The example sentences / phrases are mainly picked up from review articles (Friederici & Weissenborn, 2007; Molinaro, Barber, & Carreiras, 2011) and complemented by some other studies (Friederici, Pfeifer, & Hahne, 1993; Frisch & Schlesewsky, 2001; Kotz, Frisch, von Cramon, & Friederici, 2003; Mueller, Hahne, Fujii, & Friederici, 2005; Newman, Pancheva, Ozawa, Neville, & Ullman, 2001). The examples are taken from experiments using either auditory or visual presentation of the sentences in English, German, or Japanese. The English translations in the brackets are literal translations of the original sentences.

Violation types	Examples
Word category errors / Phrase structure violations	*Max's of proof vs. Max's proof
	*your write vs. you write
	*I believe him is a spy. vs. I believe he is a spy.
	*Die Pizza wurde im _ gegessen. (*The pizza was in-the eaten.) vs. Die Pizza wurde gegessen. (The pizza was eaten.)
	*Yesterday I cut Max's with apple caution. vs. Yesterday I cut Max's apple with caution.
	*Ichi wa no kamo ga ni hiki no _ tobikoeru tokoro desu. (One duck two _ jump over take place.) vs. Ichi wa no kamo ga ni hiki no neko o tobikoeru tokoro desu. (One duck two cats jump over takes place.)
Morphosyntactic violations	*Das Parkett wurde bohnere. (*The parquet was polish.) vs. Das Parkett wurde gebohnert. (The parquet was polished.)

Agreement violations a) subject-verb b) determiner-noun	a) *As a turtle grows its shell grow too. vs. As a turtle grows its shell grows too.
	b) *Sie bereist den [M] Land [N] auf einem kräftigen Kamel. (*She travels the [M] land [N] on a strong camel.) vs. Sie bereist das [N] Land [N] auf einem kräftigen Kamel. (She travels the [N] land [N] on a strong camel.)
Violations of verb-argument structure	*Paul fragt sich, welcher Angler [NOM] der Jäger [NOM] gelobt hat. (*Paul asks himself which angler [NOM] the hunter [NOM] praised has.) vs. Paul fragt sich, welchen Angler [ACC] der Jäger [NOM] gelobt hat. (Paul asks himself which angler [ACC] the hunter [NOM] praised has.)
	*Ichi wa no kamo ga [NOM] ni hiki no neko ga [NOM] tobikoeru tokoro desu. (One duck [NOM] two cats [NOM] jump over takes place.) vs. Ichi wa no kamo ga [NOM] ni hiki no neko o [ACC] tobikoeru tokoro desu. (One duck [NOM] two cats [ACC] jump over takes place.)
	*Das Zimmer wurde gearbeitet. (*The room was worked.) vs. Im Zimmer wurde gearbeitet. (In the room it was worked.)

In an experimental study, behavioral and/or brain responses to correct and incorrect sentences are contrasted. Existence of differences between those two responses is interpreted as an evidence for participants' ability to apply the grammar of the given language. Major tasks of those experiments include 1) passive listening/reading, 2) comprehension questions, 3) acceptability judgement, and 4) grammaticality[25] judgement (Molinaro et al., 2011). The former tasks, i.e. passive listening/reading and comprehension questions, are indirect tasks, in which the violations are not task-relevant, whereas the latter, i.e. acceptability and grammaticality judgement, are direct task, in which participants have to evaluate the sentences explicitly and thus the violations play an immediate role for the task.

[25] In the case of "grammaticality judgement", "grammaticality" is used as a synonym of "acceptability". That is, "grammaticality" is not used as a part of the formal language theory.

Moreover, violation approach comprises not only investigations of natural language sentence processing, but also those examining the processing of Jabberwocky sentences (see Table 9.2) and artificial sequences built on the basis of artificial grammars (see

Table *9.3*). Jabberwocky sentences contain morphosyntactic information bootstrapping sentence processing, but are lacking lexical semantic information because most words, except some function words, are replaced by pseudo-words. While Jabberwocky sentences rely on morphosyntactic and morphophonological rules of natural languages, artificial sequences require learning of a set of unfamiliar artificial rules (see next section). After mastering a particular artificial grammar, participants are able to differentiate sequences created by the learned grammar (correct) or not (incorrect). Processing of Jabberwocky sentences and artificial sequences are regarded as 'pure syntactic' processing without (lexical) meaning.

Table 9.2 Violation types and example Jabberwocky sentences. The example sentences are adopted from studies conducted by Münte, Matzke, and Johannes (1997), Hahne and Jescheniak (2001), Yamada and Neville (2007) and a review article by Canseco-Gonzalez (2000).

Violation types	Examples
Word category errors / Phrase structure violations	*Minno can kogg the mibe with her that nove. vs. Minno can kogg the mibe with that nove.
	*Das Fiehm wurde im _ gerottert. (*The ploker was being in-the rished.) vs. Die Glabbe wurde gerottert. (The wibon was being rished.)
	*The celtron resented Tom's about malwars her nabs and her cesting. vs. The celtron resented Tom's malwars about her nabs and her cesting.
Agreement violations subject-verb determiner-noun	a) *Das Klenck frunen den Wech. (*A mizzel quanch the plurr.) vs. Der Kruke plötzt den Schruck. (A flurk nerches the minch.) a) *Manche Verzinker trögelt den Blotz. (*Some globbies biggles the vinch.) vs. Viele Wenken donzen den Tend. (Many fluzzies brin the chink.)

Table 9.3 Artificial sequence characterizations and examples.

Sequence characterizations	Examples
(AB)n	*ABAA or *ABABABAA vs. ABAB or ABABAB
AnBn	*AABA or *AAAABBBA vs. AABB or AAAABBBB

9.1.1.2 Artificial grammar learning

Artificial grammar learning (AGL) is an experimental paradigm used in several research areas such as psychology, neuroscience, and cognitive biology to investigate the capacity to generalize over a set of strings generated by an artificial grammar (for reviews, see Fitch & Friederici, 2012; Fitch, Friederici, & Hagoort, 2012; Jäger & Rogers, 2012). A typical AGL experiment consists of two phases: a training phase and a test phase. In the training phase, the participants are familiarized with a subset of strings generated by a particular artificial grammar. The experimental design of this training phase is variable. In some cases, the participants are instructed to write or type the presented strings as accurate as possible without saying that a particular rule underlies those strings. In other cases, the subjects are explicitly asked to extract the rule underlying the presented strings. In the test phase, a discrimination task is conducted in which the subjects are required to identify whether novel strings presented in the test phase are familiar or surprising.

Many recent AGL experiments resort to formal language theory (see Section 8.1, p. 46) in generating experimental stimuli. Those experiments often employ the patterns of the types $(AB)^n$, i.e. {AB, ABAB, ABABAB, ...}, and A^nB^n, i.e. {AB, AABB, AAABBB, ...}. The former patterns, i.e. repeated AB pairs, can be generated by a regular grammar (Fitch & Friederici, 2012; Fitch et al., 2012), or rather, they belong to the simplest class called the strictly k-local language (Jäger & Rogers, 2012). The latter patterns, i.e., a number of As is followed by exactly the same number of Bs, is context-free (Fitch & Friederici, 2012; Fitch et al., 2012; Jäger & Rogers, 2012). The main interest of such empirical studies integrating AGL and formal language theory lies in testing the *supra-regular distinctiveness hypothesis* stating that "humans are unusual (or perhaps unique) among animals in processing supra-regular processing power" and the *supra-regular neural distinctiveness hypothesis* holding that "the human capacity for supra-regular grammars may be implemented in brain regions separate from those involved in simple sequencing at the sub-regular level" (Fitch et al., 2012, p. 1929).

In early neuroscientific AGL experiments (e.g., Bahlmann, Gunter, & Friederici, 2006; Fitch & Hauser, 2004; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006), A^nB^n type sequences were interpreted as center-embedding on the basis of rewrite rules introduced in the formal language theory. Indeed, application of the context-free grammar including 1) S -> aSb and 2) S -> ab creates sequences with center-embedding structure such as [a[ab]b] and [a[a[ab]b]b]. However, in these experiments, it was not clear whether participants really applied such a grammar and processed center-embedding, they processed cross-serial dependency, or they counted the number of "a" and "b" to evaluate whether b was presented as much as "a" was (for discussions, see also Fitch & Friederici, 2012; Ojima & Okanoya, 2014). To ensure that participants also process dependencies, later AGL experiments introduced a paradigm in which relationship between a and b was made explicit by using [a1[a2[a3b3]b2]b1] (e.g., Bahlmann, Schubotz, & Friederici, 2008). In this way, participants had to learn and recognize center-embedding pair-wise dependencies which are comparable to those in natural language sentences.

9.1.1.3 Isolating syntactic processing

Comparison of sentences, syntactic prose, Jabberwocky sentences, and nonstructured word lists is conducted to isolate syntactic processing (Table 9.4). In the case of the comparison between sentences and nonstructured word lists, syntactic operations such as structure building, agreement checking, and thematic-role mapping can be isolated because the latter condition requires them not at all or less[26] if any (Kaan & Swaab, 2002). Nonstructured word lists contain either only content words without function words or both, and also vary in semantic relatedness between words. Syntactic prose uses existing words to construct a sentence which is grammatically correct, but nonsense. Thus, it reduces semantic information from sentences. Jabberwocky sentences are also used to reduce semantic processing from sentence processing by replacing words through pseudo words which follow phonological rules of the language, but doesn't contain meaning (Kaan & Swaab, 2002). Because Jabberwocky sentences still contain morphosyntactic markers such as inflection and function words such as "a" and "the", structure building operation and agreement checking are recruited.

[26] Because of some syntactically legal combinations of two or three words in the word list (e.g., "Der Koch" ("The cook") in the Table 9.4), local structure building might be engaged also in processing word lists (Friederici, 2011).

Table 9.4 Examples ordered according to the strength of semantic and syntactic information involved in the sequence. The examples stem from Friederici, Meyer, & von Cramon (2000), Humphries, Binder, Medler, & Liebenthal (2006), and Kaan and Swaab (2002).

Types	Examples
Sentence	Die hungrige Katze jagt die flinke Maus. (The hungry cat chased the fast mouse.)
	The man on a vacation lost a bag and a wallet.
Syntactic prose	The infuriated water grabbed the justified dream.
	The freeway on a pie watched a house and a window.
Jabberwocky sentence	Das mumpfige Fölöfel föngert das apoldige Trekon. (The mumphy folofel fonged the apole trecon.)
	The solims on a sonting grilloted a yome and a sovir.
Semantically related word list	on vacation lost then a and bag wallet man then a
Semantically random word list	a ball the a the spilled librarian in sign through fire
	Der Koch stumm Kater Geschwindigkeit doch Ehre. (The cook silent cat velocity yet honor.)
Pseudo word list	rooned the sif into lilf the and the foig aurene to
	Der Norp Burch Orlont Kinker Deftei Glauch Leigerei (the norp burch orlont kinker deftey glaunch legery)

9.1.1.4 Manipulation of syntactic complexity

There are mainly three ways to investigate the effect of syntactic complexity (see also Table 9.5). The first one is comparison between canonical and non-canonical sentences (Friederici, 2011; Kaan & Swaab, 2002). In English, for example, canonical sentences are active sentences or sentences with subject-relative clause, while non-canonical sentences include passive sentences and sentences with object-relative clause. The second one is contrasting nesting condition with non-nesting condition (Friederici, 2011). In the nesting condition, for example, further relative clauses are embedded in a relative clause. The more relative clauses are nested, the more complex the sentence. Moreover, syntactic ambiguity is sometimes regarded as making sentences complex (Friederici, 2011). Syntactically ambiguous sentences are, for example, caused by ambiguity in syntactic category (Stowe, Paans, Wijers, & Zwarts, 2004; Tyler & Marslen-Wilson, 2008) or referential ambiguity (Sambin et al., 2012).

Table 9.5 Examples are based on Friederici (2011), Kaan and Swaab (2002), Sambin et al. (2012), and Stowe et al. (2004). The English translations in the brackets are literal translations of the original sentences.

Types	Examples
Canonical vs. non-canonical sentences	The reporter who attacked the senator admitted the error. vs. The reporter who the senator attacked admitted the error.
Non-nesting vs. nesting	Peter wusste, dass Achim (S1) den großen Mann gestern am späten Abend gesehen hatte (V1). (Achim the tall man yesterday at late night saw.) vs. Peter wusste, dass Maria (S1), die Hans (S2), der (S3) gut aussah (V3), liebte (V2), Johan geküsst hatte (V1). (Maria who Hans who was good looking loved Johann kissed.)
Syntactic ambiguous sentence	Paul smiled when he entered. ("he" can be interpreted as co-referential (he = Paul) or non-coreferential (he ≠ Paul) and thus ambiguous.)
	The red drops from the dye bottle onto the floor. ("drops" can be noun or verb at the first moment.)

9.1.2 Time course and functional neuroanatomy

Concerning the time course of auditory sentence processing, Friederici (2002) introduced a three-phase model of language comprehension reflected in different components in the event-related potential (ERP) (for details and up-dates, see also Friederici, 2011). The first phase deals with the initial build-up of local phrase structure on the basis of the syntactic category of a word. Based on ERP studies investigating brain response to phrase structure violation (early left anterior negativity; ELAN), this processing phase was suggested to take place between 100 and 300 ms. This first phase is claimed to be autonomous or modular and precedes semantic processing.

The second phase which is suggested to occur between 300 and 500 ms is about computation of syntactic and semantic relations with the goal of thematic role assignment, i.e. understanding who is doing what to whom. Processing of syntactic relations is considered to be reflected in the left anterior negativity (LAN), an ERP component elicited by

morphosyntactic violations between 100 and 500 ms, whereas processing of semantic relations is suggested to be reflected in a centroparietal negativity between 300 and 500 ms (N400) elicited by lexical-semantic violations. In this second phase, morphosyntactic and lexical-semantic processing are suggested to take place in parallel without interaction.

The third phase in the time window between 500-1000 ms concerns sentence-level integration of different types of information (e.g., syntactic and semantic information). A late centroparietal positivity (P600) regarded as reflecting this third processing phase is elicited by outright syntactic violations, 'garden-path' sentences requiring syntactic reanalysis, and syntactically complex sentences. In this late phase, syntactic and semantic processing interact.

Friederici's model of functional neuroanatomy of auditory sentence comprehension centers on dorsal and ventral language pathways as well as brain regions connected by those pathways, especially Broca's region (Friederici, 2011, 2012, 2015, 2017; Friederici & Gierhan, 2013). Language-related dorsal pathways include 1) D1 (a part of the superior longitudinal fasciculus (SLF)) connecting the temporal cortex to the dorsal premotor cortex (dPMC) via parietal cortex and 2) D2 (the arcuate fasciculus (AF); or sometimes also AF/SLF) connecting the temporal cortex to the posterior portion of Broca's region (BA 44). Language-related ventral pathways comprise 1) V1 (the extreme fiber capsule system (EFCS); or sometime also longitudinal inferior-fronto-occipital fasciculus (IFOF)) connecting the anterior portion of Broca's region (BA 45) and BA 47 to the superior temporal gyrus (STG) and 2) V2 (the uncinate fasciculus (UF)) connecting the most ventral parts of the inferior frontal region including the frontal operculum (FOP) to the temporal cortex. These four language-related cortical pathways are represented in Figure 9.1.

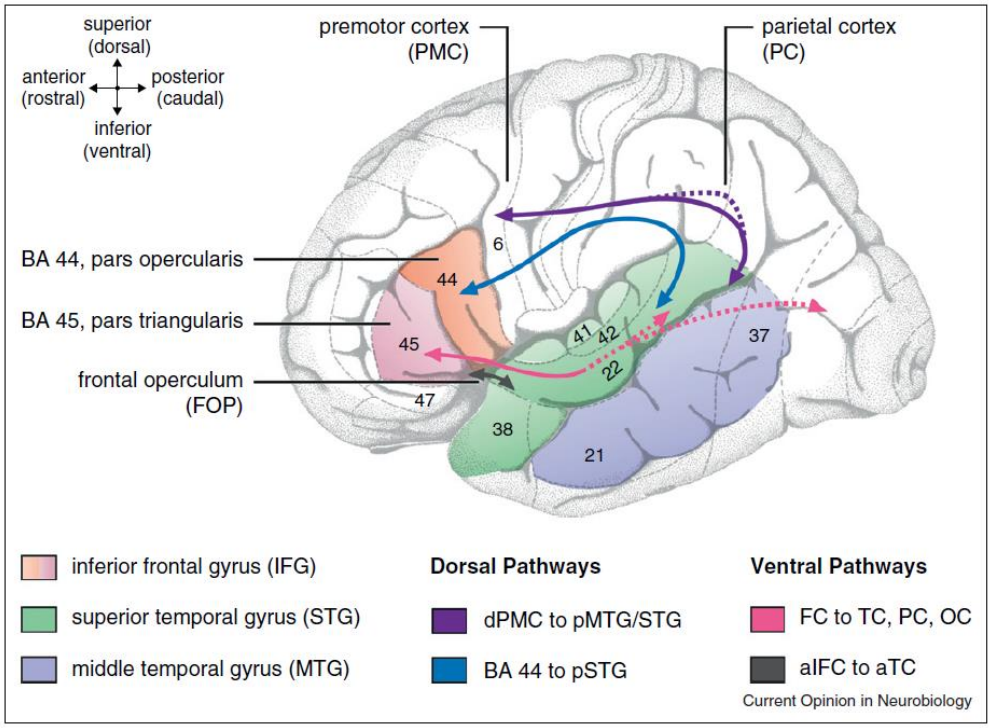


Figure 9.1 The figure is adopted from Friederici & Gierhan (2013, p. 251), *Current Opinion in Neurobiology*, 23: 250–254, with permission by Elsevier.

Concerning sentence comprehension, three of those four pathways (D2, V1, and V2) were suggested to be of particular importance (Friederici, 2015). The recent model (Friederici, 2015) especially focuses on the 'core functions' of language, including semantic and syntactic processes, and thus leaves the sensory-motor interface suggested to be supported by D1 (a part of the phonological network) among others less discussed (for detailed discussion, see Friederici, 2011). As for syntactic processing, two processing stages implemented by different syntactic networks are suggested: 1) Step 1 local structure building by ventral syntactic network (V2); and 2) Step 2 processing global, hierarchical structural dependencies by dorsal syntactic network (D2). The V1, then, is rather involved in semantic processing where controlled semantic processing reflected in N400 takes place in BA 45/47 (Friederici, 2012). The fast lexical semantic processing is located in the middle temporal gyrus (MTG).

Friederici (2011) introduced a neurocognitive model integrating the temporo-frontal network model with the time course of auditory language comprehension (Figure 9.2). Importantly, syntactic networks (pink color) are left lateralized. At step 1, the local phrase structure is automatically built on the basis of word category information within the UF connecting anterior superior temporal gyrus/sulcus (aSTG/STS) and FOP. This step, therefore, corresponds to the first phase of the three-phase model reflected by ELAN and takes place

between 100 and 300 ms. Step 2 is concerned with processing syntactically complex sentences where the posterior Broca's region (BA 44) subserves building up of non-adjacent hierarchical structure and the posterior temporal gyrus/sulcus (pSTG/STS) integration of syntactic and semantic information in complex sentences. Although it was not made explicitly, this second stage seems to subsume the second and third phases of the three-phase model introduced above.

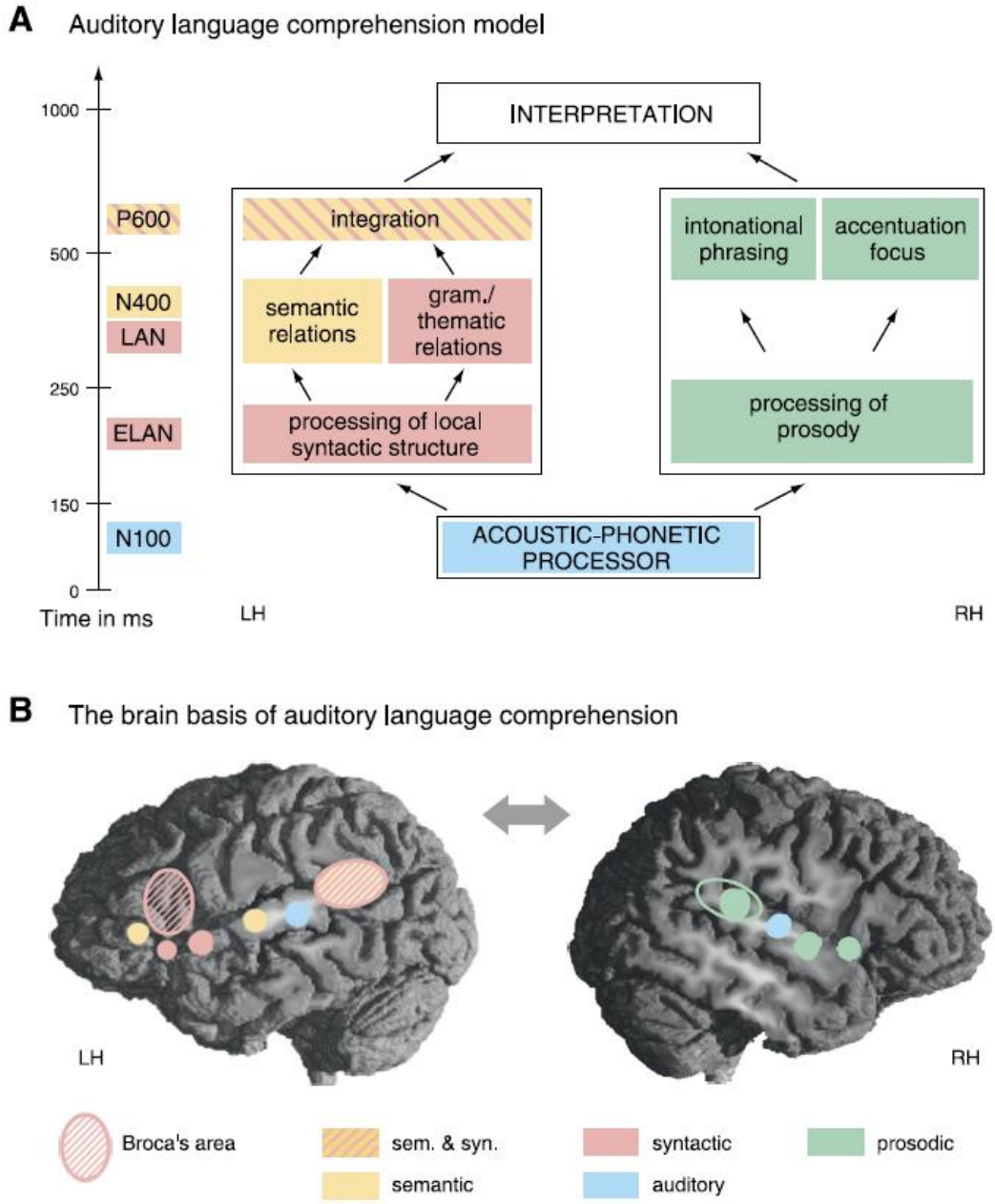


Figure 9.2 Time course and functional neuroanatomy of auditory sentence processing. The oval called Broca's area is which is called "Broca's region" in the current thesis and refers only to the region in the left IFG. The figure is adopted from Friederici (2011, p. 1377), *Physiol Rev*, 91: 1357–1392, with permission by Angela Friederici.

Friederici (2017) introduced an updated model with some changes. First, the component "processing of local syntactic structure" in Figure 9.2A turned to "syntactic phrase structure building". Second, the component "gram./thematic relations" (i.e., grammatical / thematic relations) in the Figure 9.2A is separated into two components: "thematic relations" and "syntactic relations". The former component is represented as an overlap between semantic and syntactic process in the second phase. The latter component is illustrated as purely syntactic in the second phase. Third, concerning the Figure 9.2B, Broca's region (the left BA 44) is assigned with "complex syntax". Further, "semantic" is located in the left BA 45 and 47 as well as in the region around the left anterior superior temporal sulcus (aSTS).

9.1.3 Syntactic processing within the temporo-frontal networks

As reviewed above, Friederici's model of auditory sentence comprehension assumes two types of syntactic processes implemented differently within the left temporo-frontal networks: 1) Local phrase structure building in the aSTG and FOP via UF, and 2) Processing syntactically complex sentences in the posterior Broca's region and the pSTG/STS via AF. In line with this model, two recent activation likelihood estimation (ALE) meta-analysis studies report consistent activity of the left posterior Broca's region (BA 44) and the left pSTG/STS in processing syntactically demanding sentences including noncanonical, embedded, and ambiguous sentences (L. Meyer & Friederici, 2015; Rodd, Vitello, Woollams, & Adank, 2015). Another ALE meta-analysis study, however, showed that those two regions and the dorsal pathway connecting them are reliably activated in structure building during sentence comprehension regardless of the complexity of sentences (Zaccarella, Schell, & Friederici, 2017). Moreover, the left BA 44 was reported to be consistently activated throughout the studies investigating semantically demanding sentences including semantic ambiguity, anomaly, and unrelatedness (Rodd et al., 2015). They also showed that the cluster for syntax and semantics in the left BA 44 partially overlap and challenge the claim that the left BA 44 is selectively activated in syntactic processing. None of those three studies show reliable activation of aSTG/STS for syntactic processing.

Another challenge which Friederici's model encounters is the relationship between working memory and syntactic mechanism in the Broca's region. For example, syntactically demanding object wh-questions[27] require more working memory resource than subject wh-

[27] An example of object wh-questions is "Who did the reporter attack?" (notated as "Who did the reporter attack *t*?"). In this sentence, an element was 'moved' from the last position (marked with "*t*" standing for "trace") in the

questions. In an ERP experiment, Fiebach, Schlesewsky, and Friederici (2002) varied the distance between filler and gap in subject and object wh-questions. They found that only object wh-questions showed a sustained left-anterior negativity and the amplitude of this negativity was larger when the filler-gap distance was larger. Those results were interpreted as effects of limited working memory resource. In addition, an functional magnetic resonance imaging (fMRI) experiment conducted by Fiebach, Schlesewsky, Lohmann, von Cramon, and Friederici (2005) showed significant activation of the Broca's region only for long object wh-questions (contrasted with short object wh-questions). Thus, they interpreted the role of the Broca's region in terms of working memory resource. Similarly, another fMRI experiment showed the main effect of the filler-gab distance in the left inferior frontal gyrus including the pars opercularis (BA 44) and the posterior portion of the pars triangularis (BA 45) as well as the inferior frontal sulcus (Makuuchi, Grodzinsky, Amunts, Santi, & Friederici, 2013).

The generative neurolinguistic model of sentence processing, however, emphasizes 'purely' syntactic mechanisms which are independent of other aspects of cognition such as working memory (for discussions, see also L. Meyer & Friederici, 2015). For example, Makuuchi, Bahlmann, Anwander, and Friederici (2009) conducted an fMRI experiment which showed the main effect of syntactic structure (i.e., multiple levels of embedding) and of working memory load (i.e., distance between dependent elements) in different area in the left inferior frontal region, with the former in the lateral convexity of the left pars opercularis (BA 44) and the latter in the left inferior sulcus (Figure 9.3). In those experiments, "working memory" is used to denote "maintenance cost of the verbal information for a certain period" (Makuuchi et al., 2009, p. 8363).

---

sentence to the sentence initial position (realized as "who"). As the distance between the trace and the realization is large, processing of object wh-question is considered as demanding. In comparison to object wh-questions, subject wh-questions such as "Who did attack the reporter?" (notated as "Who *t* did attack the reporter?") are regarded as less demanding to process as the trace and the realization are close to each other.

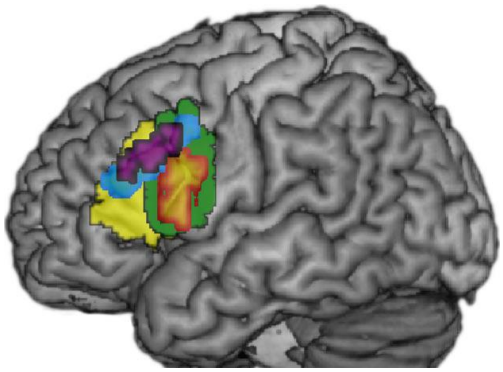


Figure 9.3 The main effect of syntactic structure (orange) and of working memory load (blue) in the inferior frontal region. The purple area showed a higher coupling with the left pars opercularis during sentences with multiple levels of embedding (in comparison with sentences with no embedding). The green area is BA 44 and the yellow area BA 45. The figure is adopted from Makuuchi et al. (2009, p. 8364), *PNAS*, 106 (20): 8362–8367, with permission by Michiru Makuuchi.

9.2 Cortical circuit model of tonal-harmonic syntax in music

9.2.1 Stefan Koelsch' neurocognitive model of music perception

A neurocognitive model of music perception which shows many parallels to Friederici's model was introduced by Stefan Koelsch (2011a). His model (Figure 9.4) consists of step-by-step processes including basic auditory information processing (Feature Extraction I and II), formation of basic musical representations (auditory sensory memory, Gestalt formation, and analysis of intervals), and music syntactic processing (structure building and structural reanalysis & repair). Musical meaning is suggested to emerge through the interpretation of (musical) information obtained through the above mentioned processing steps. Like Friederici's model, Koelsch' model is based on several neuroimaging studies investigating the time course of processing (by using electroencephalography (EEG) and magnetoencephalography (MEG)) as well as brain regions activated by those processes (by using fMRI). Below, hypothesized neural correlates of structure building components in music syntactic processing are amplified and discussed.

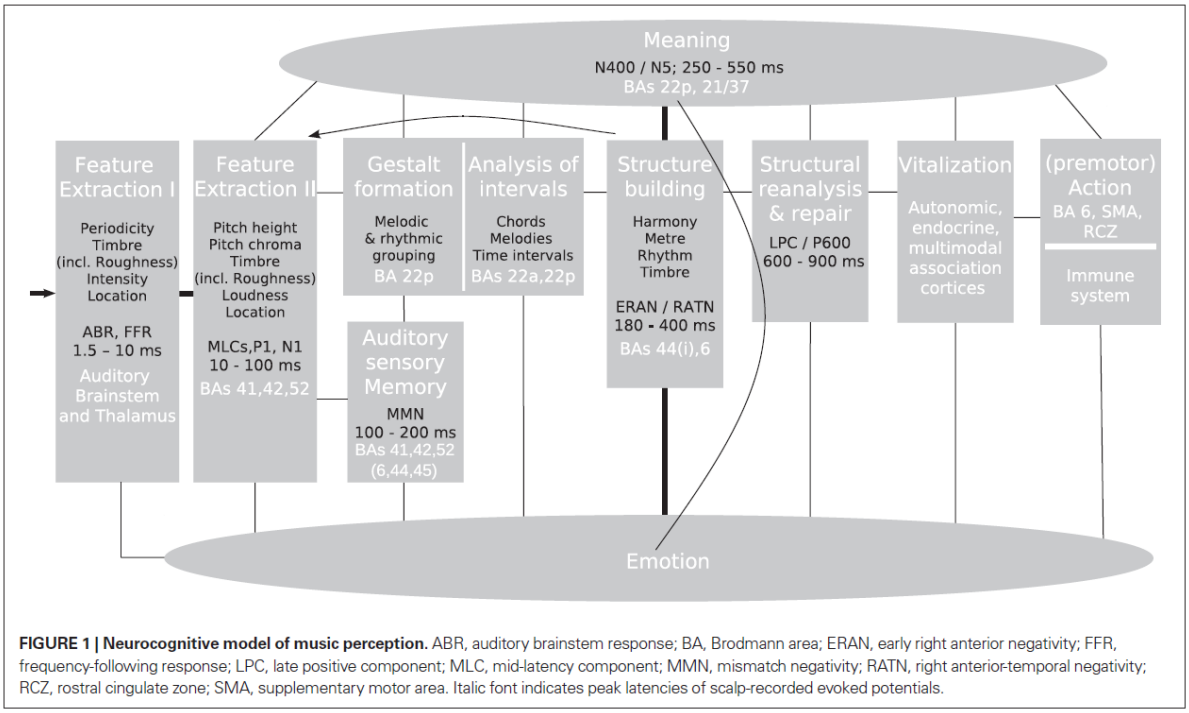


FIGURE 1 | Neurocognitive model of music perception. ABR, auditory brainstem response; BA, Brodmann area; ERAN, early right anterior negativity; FFR, frequency-following response; LPC, late positive component; MLC, mid-latency component; MMN, mismatch negativity; RATN, right anterior-temporal negativity; RCZ, rostral cingulate zone; SMA, supplementary motor area. Italic font indicates peak latencies of scalp-recorded evoked potentials.

Figure 9.4 Neurocognitive model of music perception. The figure is adopted from Koelsch (2011a, p. 3), *Frontiers in Psychology*, 2: 110.

Music syntactic processing studied by Koelsch and his colleagues refers to "the processing of major-minor tonal syntax, particularly with regard to chord functions (i.e., with regard to harmony [...])" (Koelsch, 2011a, p. 5). Therefore, I use the term "tonal-harmonic syntactic processing" below instead of "music syntactic processing" which includes other domains of musical syntax. According to Koelsch (2012a, 2012b), tonal-harmonic syntactic processing consist of several sub-processes such as extracting tonal center, establishing structural relationships between pitch-events (e.g., stability relationship and long-distance dependency), hierarchical structure building as well as structural reanalysis and repair. In addition, he suggested that further general processes such as prediction, implicit learning, working memory, and attention are involved in tonal-harmonic syntactic processing. Most of the studies conducted by Koelsch and colleagues which investigate tonal-harmonic syntactic processing used the violation paradigm to (indirectly) study structure building by mainly using EEG and fMRI.

One electrophysiological correlate of tonal-harmonic syntactic processing is the early right anterior negativity (ERAN), i.e. an event-related potential (ERP) with negative polarity, maximal amplitude values at (right) frontal electrodes, and a peak latency around 150-180 ms (Figure 9.5C) (Koelsch, 2011a, 2012b). This brain potential was first discovered in an

experiment using a chord sequence paradigm (Figure 9.5B), in which ERAN was elicited by irregular chords in comparison to regular chords (Koelsch, Gunter, Friederici, & Schröger, 2000). Afterwards, ERAN was repeatedly found in similar experiments using a chord sequence paradigm with partly modified stimuli and/or tasks as well as extended experiments using polyphonic stimuli, existing musical piece, and melody (for reviews, see Koelsch, 2009, 2011a, 2012a). In all those studies, harmonic or tonal violations were considered as tonal-harmonic syntactic incongruities. ERAN was suggested to reflect a violation of harmonic structure building (Koelsch, 2012a). Moreover, cognitive processes reflected in ERAN are shared with those reflected in LAN because they show an interaction (i.e., reduced amplitude) if syntactically incongruent words and chords are simultaneously presented (Koelsch, Gunter, et al., 2005; Steinbeis & Koelsch, 2008).

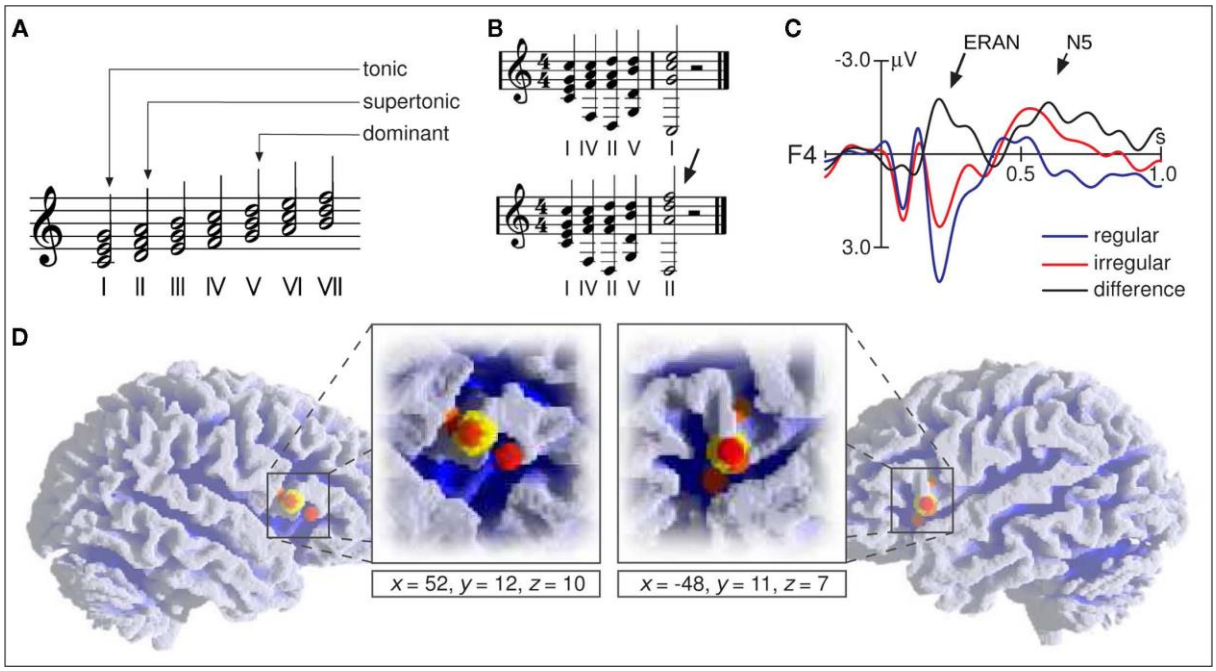


Figure 9.5 Illustration of A) chord functions, B) chord sequence paradigm, C) ERAN and N5, and D) source localization of ERAN. The figure is adopted from Koelsch (2011a, p. 5), Frontiers in Psychology, 2: 110, with permission from Elsevier for the original material published by Koelsch and Siebel (2005, p. 580), *Trends in Cognitive Sciences*, 9 (12): 578-584.

Neuroanatomical correlates of tonal-harmonic syntactic processing were suggested to be the inferior pars opercularis of the inferior fronto-lateral cortex (BA 44v) with additional contributions from the ventrolateral premotor cortex (BA 6) and the anterior superior temporal gyrus (planum polare) (Koelsch, 2012a, p. 131). These areas were more activated in the experimental condition including chord sequences with irregular, less-related or unrelated

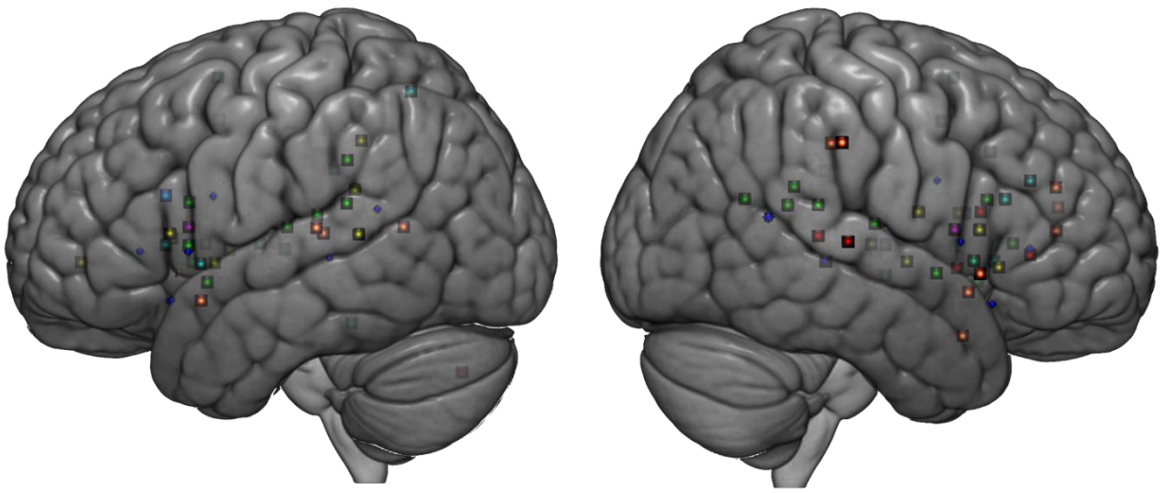
chords in comparison to regular chord sequences largely throughout different EEG and MEG source localization studies, fMRI studies, and recordings from intracranial grid-electrodes from patients with epilepsy (Bianco et al., 2016; Garza Villarreal, Brattico, Leino, Østergaard, & Vuust, 2011; Koelsch et al., 2002; Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005; Maess, Koelsch, Gunter, & Friederici, 2001; Musso et al., 2015; Sammler, 2009; Tillmann et al., 2006; Tillmann, Janata, & Bharucha, 2003; for short summary, see Table 9.6; for details see Figure 9.6). The bilateral BA 44 seem to be consistently more activated in experimental conditions including harmonic structure violations. Stronger activation (irregular (or unrelated or less-related) > regular (or related)) of the aSTG (and pSTG) was reported in some studies. Figure 9.6 also shows that peak activations in the left BA 44 are mainly at its ventral part. That is, this analysis suggests bilateral BA 44, BA 6, aSTG, and pSTG as possible candidate neural correlates of tonal-harmonic syntactic processing and underlines the role of BA 44 (but see discussions below and in Section 9.2.2, p. 75).

Table 9.6 The table summarizes the activations of the neural regions hypothesized by Koelsch (2012a) as neuroanatomical correlates of tonal-harmonic syntactic processing and a further region (pSTG) additionally reviewed in Friederici (2011) in each study. In this table, only studies using chord sequence paradigm are reviewed. Only results gained by the subtraction analysis (irregular (or unrelated or less-related) - regular (or related)) are reviewed.

	Methods (Participants)	BA44/45 (BA45)	BA44	BA44/6	BA6	aSTG/ planum polare	pSTG/ planum temporale
Maess et al. (2001)	MEG source localization (nonmusicians)	-	bilateral	-	-	-	-
Koelsch et al. (2002)	fMRI (nonmusicians)	-	bilateral	bilateral	-	bilateral	bilateral
Tillmann et al. (2003)	fMRI (adults)	bilateral	-	-	bilateral	-	-

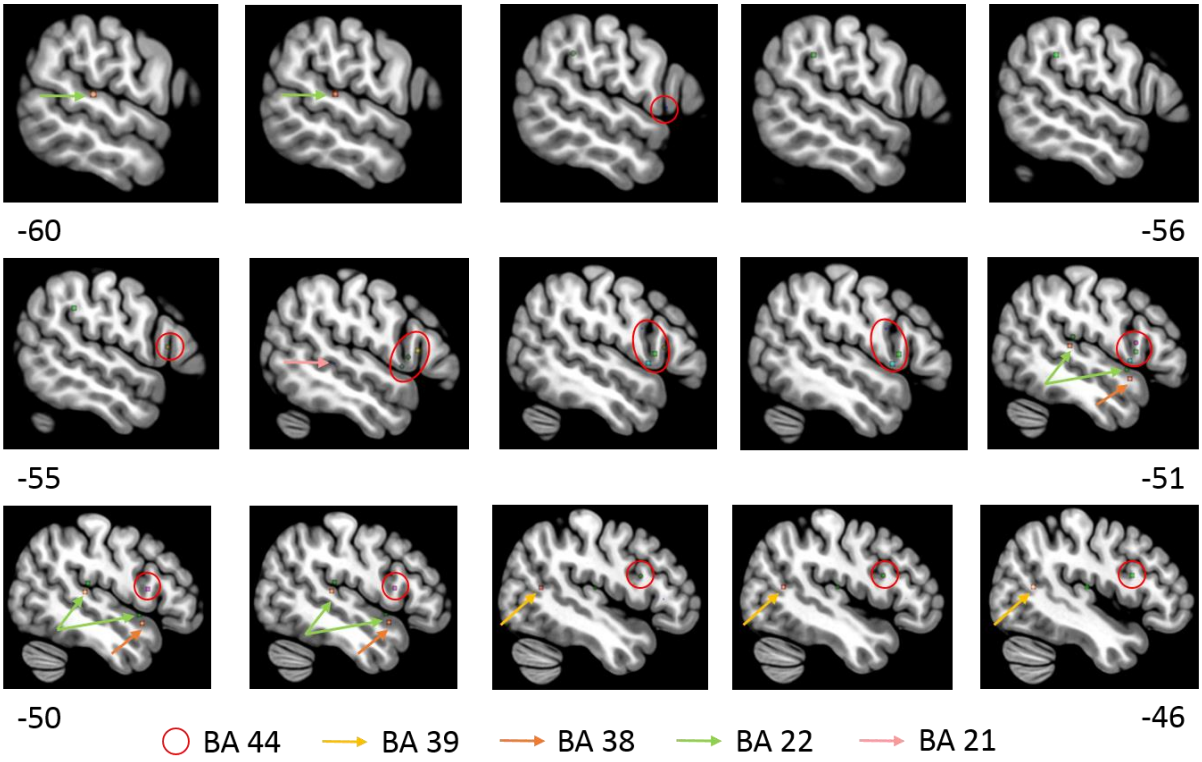
Koelsch et al. (2005)	fMRI (adults) (children)	(bilateral) (right)	bilateral right	- -	bilateral right	- -	bilateral right
Tillmann et al. (2006)	fMRI (adults)	-	right	-	-	left	left
Villarreal et al. (2011)	EEG source localization (adults)	bilateral	-	-	-	-	-
Musso et al. (2015)	fMRI (adults)	(bilateral)	-	-	-	-	-
Bianco et al. (2016)	fMRI (pianists)	(right)	-	-	-	-	right

a)



b)

Left hemisphere



Right hemisphere

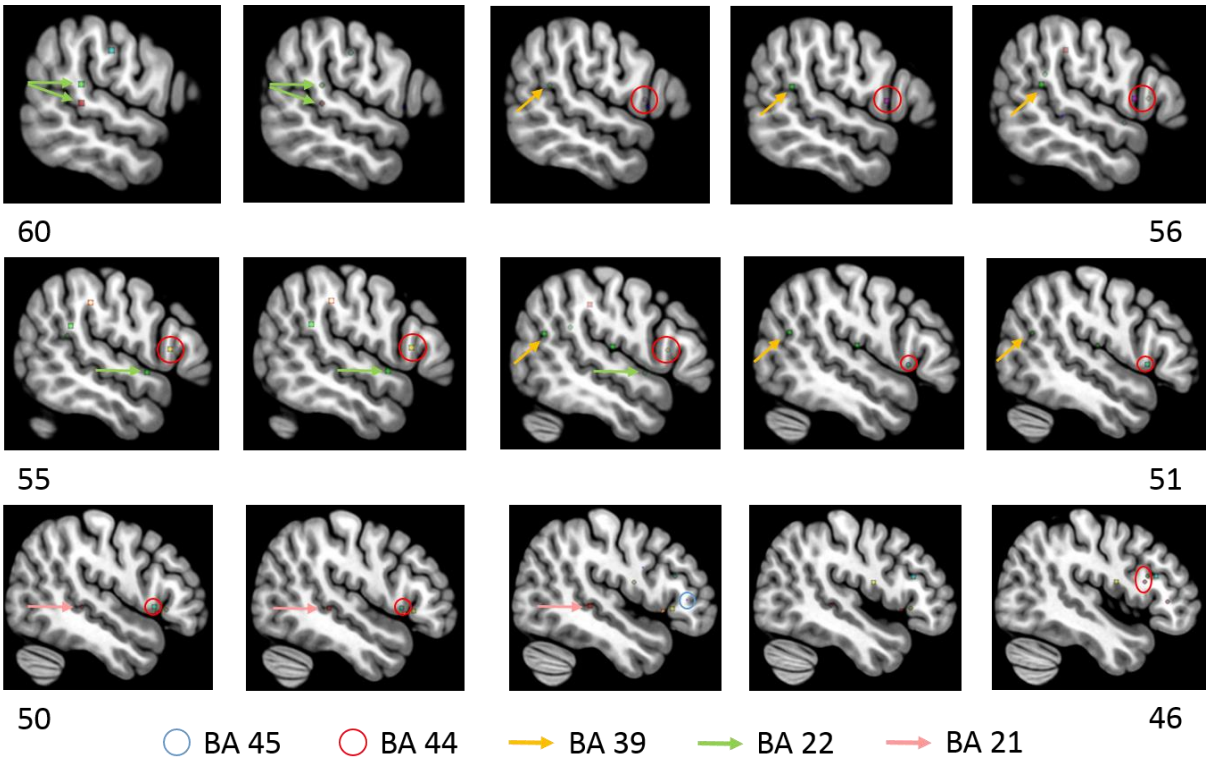


Figure 9.6 The figure is an overview of the peak activation foci (in MNI space[28]) reported in the above-mentioned studies. Only adults' data are included. If the coordinates are reported in Talairach space in the original studies, they are transformed by using convert foci function (Talairach to MNI (SPM[29])) implemented in the BrainMap GingerALE software[30]. The figures were created by MRIcroGL software[31]. a) Lateral view: The peak activations are projected onto the nearest surface. The intensity/brightness of the color represents the distance of the foci to the surface. b) Sagittal view: The peak activation foci in the Broca's region and the temporal lobe are labeled with the Brodmann areas.

However, a neuropsychological study comparing ERAN of healthy controls and patients with left IFG lesion showed difference only in the scalp distribution (ERAN was more anteriorly distributed and strongly right lateralized in the patient group) and not in the amplitude (Sammler, Koelsch, & Friederici, 2011). The only finding in favor of the claim that the left IFG is necessary for tonal-harmonic syntactic processing was a correlation between years of injury and hit rate of the regular-irregular discrimination task (the post-hoc behavioral experiment was conducted approx. two days after the main experiment session) and the chance level hit rate of this task in the patient group (even though there was no statistically significant difference to the hit rate of the control group). Moreover, this study also examined patients with the aSTG lesion and showed that there was no difference between control and patient group. Therefore, the necessity of the Broca's region and the aSTG for tonal-harmonic syntactic processing is still unclear.

9.2.2 Neural correlates of tonal-harmonic syntactic processing: An ALE meta-analysis

As reviewed above, neuroimaging studies using chord-sequence violation paradigm provided evidences for engagement of the Broca's region (especially BA 44) and BA 6 as well as the anterior and posterior STG in tonal-harmonic processing. However, it is not clear whether they are consistently involved in an extended range of neuroimaging studies

[28] MNI (Montreal Neurological Institute) space and Talairach space are based on different standardized 3D coordinate frames.
[29] SPM (statistical parametric mapping) is a statistical analysis technique used frequently in fMRI data analysis. SPM (https://www.fil.ion.ucl.ac.uk/spm/) is also name of an analysis software (available as a MATLAB toolbox) implementing this idea.
[30] GingerALE software (http://www.brainmap.org/ale/) is a software to conduct an ALE meta-analysis. It has another function for transforming the data from one coordinate system to another.
[31] MRIcroGL (http://www.cabiatl.com/mricrogl/) is a software for displaying 3D medical image.

investigating tonal-harmonic syntax. Thus, an activation likelihood estimation (ALE) meta-analysis was conducted on seventeen fMRI studies investigating any of the above-mentioned sub-processes in tonal-harmonic syntactic processing. First of all, neuroimaging studies were retrieved by using PubMed and Google Scholar with "Music AND Syntax", "Music AND Structure AND fMRI", "Music AND Grammar AND fMRI", "(fMRI OR functional magnetic resonance imaging) AND (music OR pitch OR melody OR harmony) AND (syntax OR structure)", and "fMRI AND music AND tension" as search criteria (Date: July 19, 2018). The studies not using fMRI, not focusing on tonal-harmonic syntactic processing, or not measuring healthy adults were excluded. The studies reporting only the results of the region of interest (ROI) analysis as well as stating no coordinate or coordinate space were also removed. Further, all remaining studies were evaluated whether they relate to sub-processes of tonal-harmonic syntactic processing and thus enter into the meta-analysis. The criteria designed on the basis of computational-representational theories (Section 8.2, p. 50) and Koelsch' sub-processes (Section 9.2.1, p. 69) were applied for the evaluation, which matched seventeen experiments (see Table 9.7).

Table 9.7 Summary of the studies included in the ALE meta-analysis on tonal-harmonic syntactic processing. Only the first authors' names are displayed in the table (Bianco et al., 2016; Cheung, Meyer, Friederici, & Koelsch, 2018; Durrant et al., 2007; Fedorenko, Behr, & Kanwisher, 2011; Foster & Zatorre, 2010; Fujisawa & Cook, 2011; Koelsch, Fritz, et al., 2005; Koelsch et al., 2002; Lehne, Rohrmeier, & Koelsch, 2013; Levitin, 2003; Musso et al., 2015; Oechslin, Van De Ville, Lazeyras, Hauert, & James, 2013; Schmithorst, 2005; Seger et al., 2013; Spada, Verga, Iadanza, Tettamanti, & Perani, 2014; Tillmann et al., 2003, 2006).

Sub-processes	Musical domain	Studies	
Extracting tonal center	Tonality	Durrant, 2007 (N = 16)	tonal > atonal
Establishing structural relationships & Structure building	Transposition	Foster, 2009 (N = 20)	transposed > simple
	Grammaticality, regularity, or relatedness	Tillmann, 2003 (N = 15)	related > unrelated
		Koelsch, 2005 (N = 20)	irregular > regular
		Tillmann, 2006	less related > related

		(N = 21)	
		Spada, 2014 (N = 19)	altered melody > correct
		Musso, 2015 (N = 11)	struct. dev. > well-formed
		Bianco, 2016 (N = 29)	incongruent > congruent
		Cheung, 2018 (N = 17)	ungramm. > gramm.
	Ordering of pitch events	Levitin, 2003 (N = 13)	normal > scrambled
		Schmithorst, 2005 (N = 15)	melody > random tones
		Fedorenko, 2011 (N = 12)	intact > scrambled
	Tension-relaxation pattern	Lehne, 2014 (N = 25)	tension regressor
	Cadence	Fujisawa, 2011 (N = 12)	cadence > white noise
		Oechslin, 2013 (N = 58)	transgression
		Seger, 2013 (N = 11)	cadences > baseline
Reanalysis and repair	Modulation	Koelsch, 2002 (N = 10)	modulation > in-key
N of studies = 17, N of subjects = 324, N of foci = 193			

The ALE meta-analysis was run by using BrainMap software GingerALE version 2.3.6 (http://www.brainmap.org). First, the foci reported in the Talairach space were converted into MNI space by using convert foci function (Talairach to MNI (SPM)) implemented in the GingerALE. Second, the ALE meta-analysis was performed. This is a coordinate-based analysis of the seventeen studies by using the ALE algorithm as implemented in GingerALE to identify the convergent foci over different studies (Eickhoff et al., 2009; Laird et al., 2005;

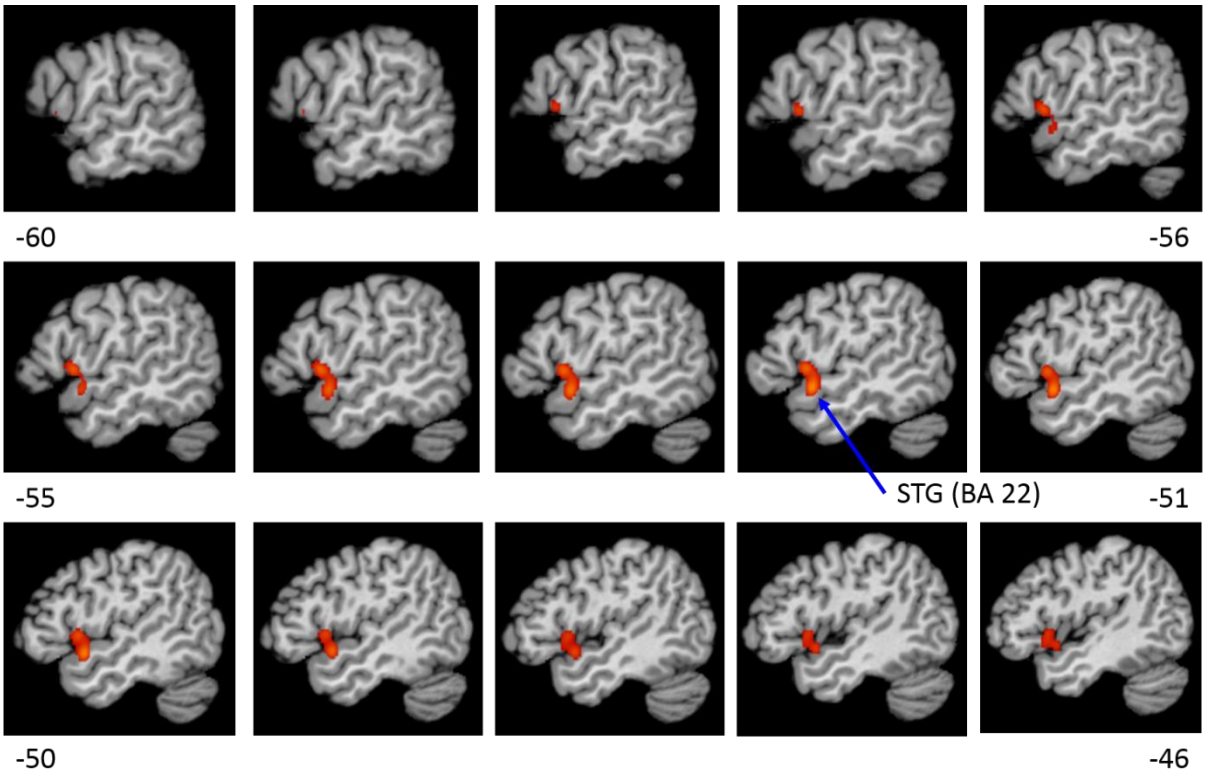
Turkeltaub, Eden, Jones, & Zeffiro, 2002). The maps were thresholded by using a cluster-level family-wise error (cFWE) correction ($P < 0.05$) with a cluster-forming threshold of $P < 0.001$ using 1,000 permutations[32]. The results of the analysis are reported in the Table 9.8. The anatomical labels were automatically generated by Talairach daemon software (Lancaster et al., 1997, 2000) included in the GingerALE[33]. Third, the ALE image was overlaid onto the MNI template (Colin27_T1_seg_MNI.nii) by using the software Mango[34]. The reported clusters and the anatomical labels of the peaks are represented in Figure 9.7.

[32] A study shows that the cluster-level FEW thresholds can be reliably estimated with 1,000 permutations (Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012).
[33] It converts MNI space to Talairach space to label anatomical regions appropriately. Thus, the outputs of the Talairach daemon software can be used as the anatomical labels of the MNI coordinates.
[34] Mango (http://rii.uthscsa.edu/mango/) is a software for displaying medical image.

Table 9.8 Results of the ALE meta-analysis on tonal-harmonic syntactic processing.

Cluster	BA	MNI coordinates (mm)			ALE (x 10^{-2})	Cluster size (mm^3)
		x	y	z		
1 (right)	13	34	24	0	2.58	7528
	44	52	18	8	2.23	
	13	48	16	-2	2.08	
	13	44	34	2	1.88	
	46	44	20	16	1.48	
	*	38	12	-2	1.38	
	47	50	22	-12	1.33	
	9	48	22	26	1.11	
2 (left)	22	-52	6	-10	1.79	2696
	22	-52	10	0	1.61	
3 (right)	22	60	-34	8	1.87	
	22	52	-32	0	1.21	
	22	58	-40	-4	1.14	

Left hemisphere



Right hemisphere

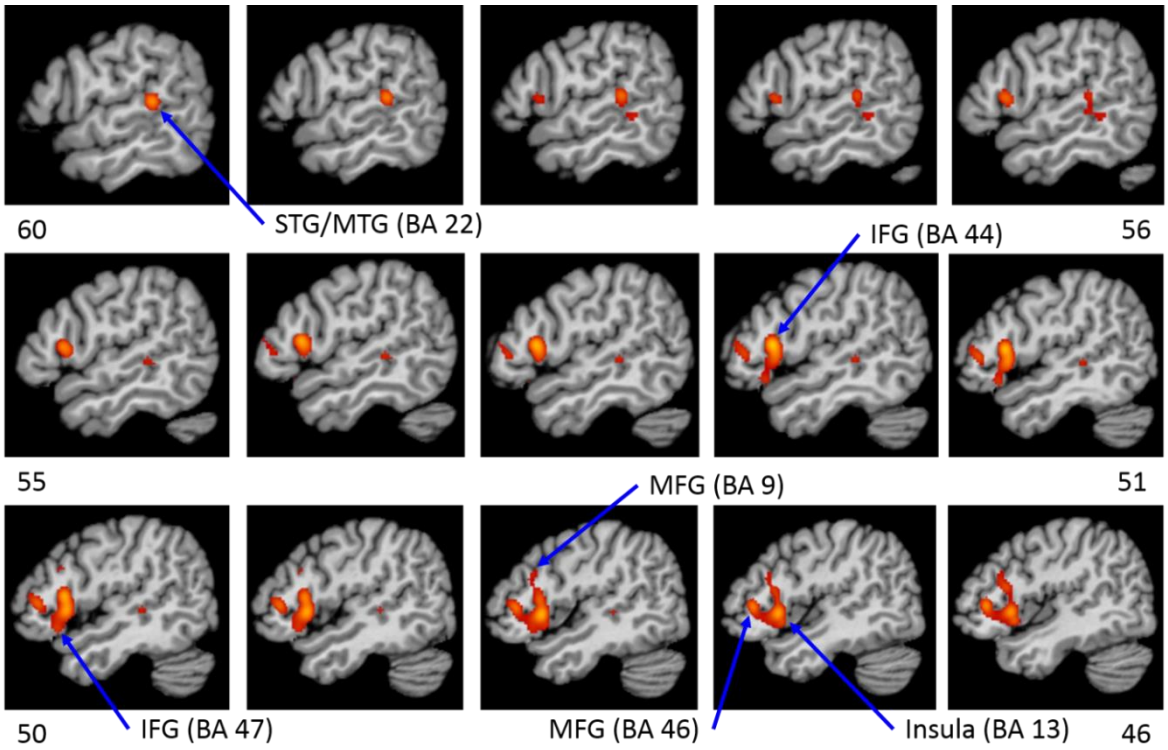


Figure 9.7 ALE image of the meta-analysis on tonal-harmonic syntactic processing overlaid onto the MNI template (Colin27_T1_seg_MNI.nii) by using the software Mango.

The analysis revealed three clusters (one large cluster and two small clusters) which achieved significance, i.e., reliable activations over different studies (Table 9.8, Figure 9.7, and Appendix A, p. 228). The first cluster is a large cluster in the right hemisphere and comprises the insula (BA 13) and the pars opercularis (BA 44) extending to the more anterior part of the inferior frontal gyrus (BA 47) and to the middle frontal gyri (BA 46 and BA 9). The studies investigating grammaticality (Bianco et al., 2016; Cheung et al., 2018; Koelsch, Fritz, et al., 2005; Spada et al., 2014; Tillmann et al., 2003, 2006), ordering of pitch events (Fedorenko et al., 2011; Levitin, 2003), cadence (Fujisawa & Cook, 2011; Oechslin et al., 2013; Seger et al., 2013), and modulation (Koelsch et al., 2002) contributed to this cluster. The second cluster is a small cluster in the left hemisphere and consists of the anterior superior temporal gyrus (BA 22). A subset of the studies listed concerning the first cluster contributed to this second small cluster (Fedorenko et al., 2011; Fujisawa & Cook, 2011; Koelsch, Fritz, et al., 2005; Koelsch et al., 2002; Levitin, 2003; Seger et al., 2013; Tillmann et al., 2003, 2006). The third cluster is a small right hemispheric cluster including the superior and middle temporal gyrus (BA 22). Only studies investigating grammaticality (Bianco et al., 2016; Cheung et al., 2018; Koelsch, Fritz, et al., 2005), ordering of pitch events (Fedorenko et al., 2011), and modulation (Koelsch et al., 2002) contributed to this cluster.

Those results indicate that the regions suggested as neuroanatomical correlates of tonal-harmonic syntactic processing by Koelsch (2012b) show reliable activation over different studies. However, there are some unique constraints. First, concerning the Broca's region, only the right pars opercularis (BA 44) reached significant ALE score. The left BA 44 is included in the nearest gray matters (within +/– 5 mm) of the cluster 2 and is involved in the dorsal end of the cluster, but it did not get a significant ALE score (see also discussions on the limitations of the current meta-analysis below). Second, the involvement of the anterior superior temporal gyrus was shown only in the left hemisphere. Third, the ventrolateral premotor cortex (BA 6) was not included in the clusters. In addition, although the left posterior superior temporal gyrus suggested to be involved in syntactic processing (Friederici, 2011) was not included in the clusters, the right posterior superior temporal gyrus was. Moreover, in addition to those regions hypothesized to be involved in tonal-harmonic syntactic processing, the insula, the middle frontal gyrus (BA 46 and 9), and another area in the inferior frontal gyrus (BA 47) showed significant ALE scores.

There are several limitations of the current meta-analysis study. First, it was not able to automatically and objectively evaluate which studies relate to sub-processes of tonal-harmonic syntactic processing and thus enter into the meta-analysis as there are no established criteria

(e.g., stimuli and tasks) associated with each sub-process. Second, although it meets the least number of studies (i.e., seventeen studies) required to gain reliable results (Eickhoff et al., 2016), clearly more studies are required. Especially, seven out of seventeen experiments examined in the current meta-analysis used violation paradigm and, thus, it is also possible that the results strongly reflect structural violation processing rather than structure building per se. Third, the current meta-analysis is a coordinate-based one and excluded experiments using ROI analysis as well as stating no coordinate or coordinate space. Some excluded studies suggest the involvement of the left inferior frontal gyrus in tonal-harmonic syntactic processing (Kunert et al., 2015; Minati et al., 2008), indicating that more studies are needed to evaluate the involvement of the left inferior frontal gyrus in tonal-harmonic syntactic processing.

In sum, the current meta-analysis shows a right hemispheric fronto-temporal network for tonal-harmonic syntactic processing with an additional area in the anterior superior temporal gyrus extending presumably to the left pars opercularis (Figure 9.8). This corresponds to the regions (including the left superior temporal gyrus, the insula, and the right inferior frontal gyrus) and the (left) temporo-insular pathway running ventrally along the extreme capsule which were suggested to rather uniquely involved in tonal-harmonic syntactic processing (Musso et al., 2015). In addition, the arcuate fasciculus connecting the right frontal and temporal area seems to be more relevant to tonal-harmonic syntactic processing.

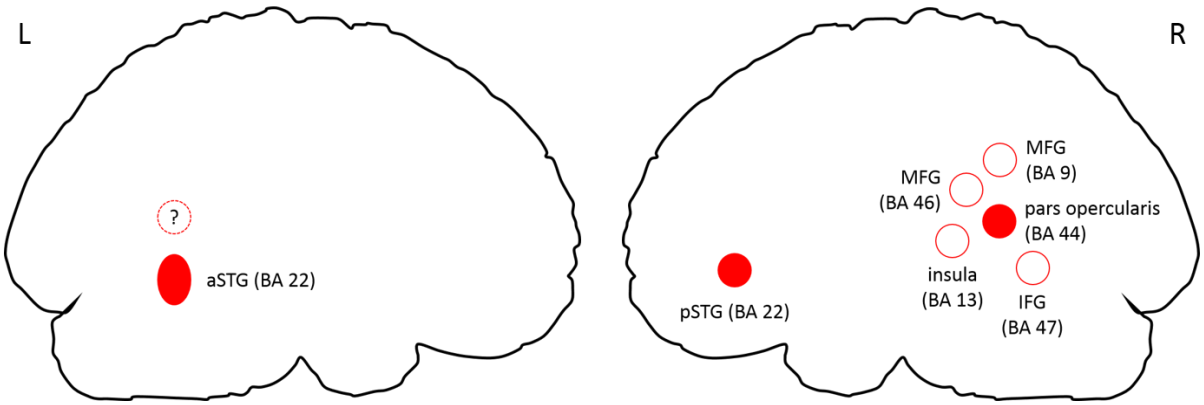


Figure 9.8 Schematic illustration of regions involved in tonal-harmonic syntactic processing. The circle with dashed line containing a question mark is BA 44. L = Left hemisphere; R = Right hemisphere

9.3 Hierarchical processing as shared aspects of linguistic and musical syntax

The investigations of the relationship between language and music from the perspective of the generative neurolinguistics and neuromusicology provided, as reviewed above, rich

comparative options. According to this approach, at the computational level, computational problems of language and music can be solved on the basis of abstract principles such as recursion and hierarchical structure building. In addition, the relationship between elements are further determined in terms of head-complement or head-elaboration. At the algorithmic and implementational level, language and music processing was investigated on the basis of syntax-centered parsing models. Especially, commonality of hierarchical processing implemented in BA 44 for both domains was pointed out repeatedly (Fitch & Martins, 2014; Jeon, 2014; Koelsch, 2011a, 2012b). In this context, hierarchical processing can be understood as assigning hierarchical structures to sequences (also called structure building) and processing (non-adjacent) structural relationships between elements. Hierarchical processing is assumed on the basis of computational-representational theories asserting that language and music are structured hierarchically. In this way, generative neurolinguistics and neuromusicology take direct correlational mapping approach (see PART I, Section 5.2.1, p. 23).

To support this top-down assumption from computational-representational theory, the argument for shared hierarchical processing in language and music from cognitive neuroscience perspectives is often made by a reverse inference (see PART I, Section 5.2.2, p. 26) as follows: if hierarchical structure is processed, Broca's region is activated; Broca's region is activated in both language and music; that is, in both language and music hierarchical structure is processed (and this process might be shared). To optimize this reverse inference, however, the selectivity of Broca's region activity for hierarchical processing should be investigated. In doing so, the Broca's region should be investigated as consisting of several subcomponents (Amunts et al., 2010). For example, Fedorenko and colleagues showed that there are (mostly) language-specific area and domain-general area side-by-side within the Broca's region (Fedorenko et al., 2011; Fedorenko, Duncan, & Kanwisher, 2012): language-specific area especially seems to lie in BA 44v and 45p. This is in line with the finding of a study using a meta-analytic connectivity modeling that the left BA 44 can be separated into five functionally separable clusters with 'language cluster' being in the anterior dorsal BA 44 and 'action and music cluster' (or 'sequencing cluster') in the posterior ventral BA 44 (Clos, Amunts, Laird, Fox, & Eickhoff, 2013). Figure 9.9 shows tonal-harmonic syntactic regions (green; the results of the ALE meta-analysis in the Section 9.2.2, p. 75) together with the regions associated with language comprehension[35] (red) retrieved from Neurosynth database

[35] Note that this is not limited to syntactic processing. By September 22, 2018, the Neurosynth did not have custom meta-analysis implemented in the online tool.

using another kind of meta-analysis (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). This also demonstrates that regions associated with language comprehension show little overlap with tonal-harmonic syntactic regions.

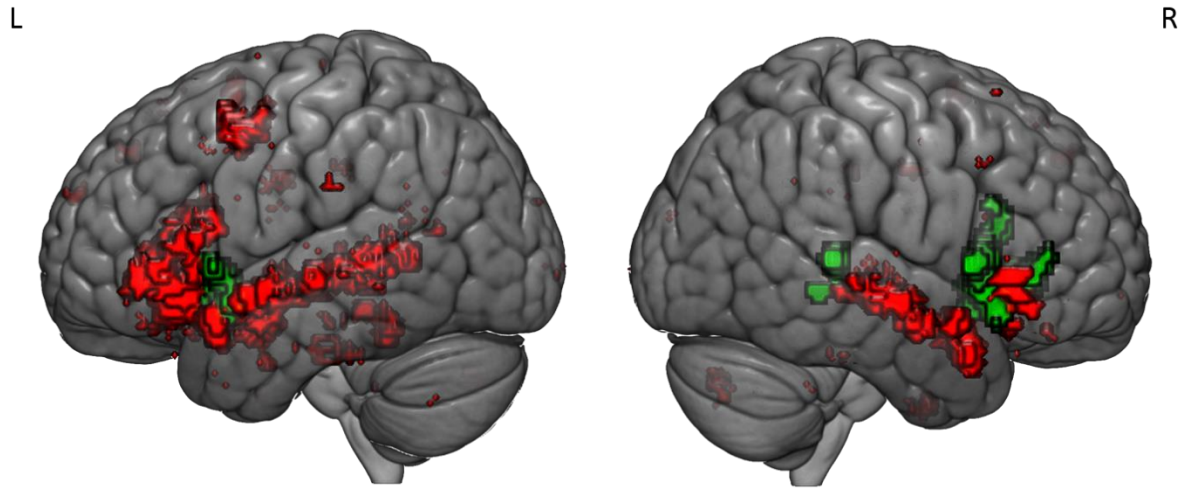


Figure 9.9 The result of the meta-analysis of musical syntax (green) in plotted together with the data for language comprehension (red) retrieved on September 22, 2018 from Neurosynth database.

Moreover, given that hierarchical processing includes two components, i.e., structure building and establishing structural relationships, those components should be examined as different concepts. For example, on one hand, structure building in terms of linguistic constituent structure building was claimed to be implemented rather in the left temporal cortex (Brennan et al., 2012; Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016). On the other hand, the effect of processing nested dependencies is located in the Broca's region (Bahlmann et al., 2008; Makuuchi et al., 2009). In both cases, it deals with hierarchical processing, but the former might be more related to structure building, while the latter to recognizing structural relationships. To investigate this issue, first of all, those two aspects of hierarchical processing should be delineated and then the probability of hierarchical structure building and recognizing structural relationships given unique tasks and stimuli should be enhanced. Another method might be one utilized by Brennan and colleagues (2016), in which different parsing models are used as predictors of brain activation patterns during (more) naturalistic language processing.

The delineation of hierarchical processing in two independent concepts such as structure building and recognizing structural relationships might be also helpful for music research as well as comparative language-music research. For example, although there are

increasing evidences that local and long-distant structural relationships are processed in music (Cheung et al., 2018; Koelsch, 2011a; Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Van de Cavey & Hartsuiker, 2016), it is still matter of debate to what extent the hierarchical structure of music determined at the computational level is built and reflected in online processing (Rohrmeier, 2011) and how musical structure building relates to that in language. In addition, linguistic syntactic ERP component LAN is elicited when processing of grammatical relations are disrupted (Friederici, 2011). Thus, shared mechanism reflected in the interaction between LAN and ERAN could be one underlying recognition of structural relationships rather than structure building as such.

10 Perspectives from neurocognitive psycholinguistics and psychomusicology

10.1 The Shared Syntactic Integration Resource Hypothesis and working memory

Dependency Locality Theory (DLT) is one of psycholinguistic theories which accounts for perceived differences in linguistic complexity as well as preference in syntactic ambiguity resolution. DLT examines these aspects in terms of "computational resources in sentence parsing that relies on [...] two kinds of resource use": "performing structural integrations" (i.e., "connecting a word into the structure for the input thus far") and "keeping the structure in memory" (i.e., "keeping track of incomplete dependencies") (Gibson, 2000, p. 95). Both structural integration processes and storage are suggested to make use of the same working memory resources (Gibson, 1998, 2000). According to DLT, processing of resource intensive sentences can be investigated in terms of the integration cost of two elements determined by the distance or locality between the two, i.e. the longer the distance between elements to be integrated (underlined below), the higher the integration cost (Gibson, 2000). Resource intensive sentences are, for example, complex sentences including object-extracted relative clauses (RC) or many layers of center embedding: the sentence (10-1) and (10-3) are much harder to process (and thus takes much longer to process) than (10-2) and (10-4) respectively (for details, see Gibson, 1998, 2000).

(10-1) The reporter who the senator attacked admitted the error. (Object-extracted RC)
(10-2) The reporter who attacked the senator admitted the error. (Subject-extracted RC)
(10-3) The reporter who the senator who John met attacked disliked the editor. (Two layers of center embedding)
(10-4) The reporter who the senator attacked disliked the editor. (One layer of center embedding)

Based on DLT and Fred Lerdahl's Tonal Pitch Space theory (TPS), Patel (2003) introduced a resource sharing framework for language and music syntactic processing. He suggested that structural integration, i.e., mentally connecting an incoming element (e.g., word or chord) to another element in the evolving structure, is a key part of syntactic processing and proposed the 'Shared Syntactic Integration Resource Hypothesis' (SSIRH) stating that language and music syntactic processing make use of the same neural resources for structural integration. In language, the integration cost is determined in terms of the distance between incoming events and their dependents. In music, it deals with the tonal distance between context chords and incoming events. The SSIRH concerns with neural resources provided by 'processing regions' in the brain which control the activation level of items in 'representation regions'. The neural resources of the processing regions are shared in language and music syntactic processing. The representation regions are understood as the loci where the structural integration itself takes place, on one hand, and the parts where domain-specific long-term memory representations are placed, on the other hand. Patel (2003) introduced the hypothesized loci of control regions (processing regions) as the frontal regions and long-term storage regions (representation regions) as the posterior (or rather temporal) regions, but not the regions where the actual integration takes place.

In a later writing, Patel (2013, p. 346) seems to equate "structural integration" with "syntactic unification" proposed to be implemented in BA 44 and 45 (Hagoort, 2005, 2013, 2016). Patel (2013) seems to see similarity between the main ideas of SSIRH and Peter Hagoort's memory unification control (MUC) model (Hagoort, 2016) because both suggest a domain-general processing component for structural integration as implemented in the frontal cortex and a domain-specific memory representation in the temporal cortex. Although further details are not discussed by Patel (2013), I elaborate this discussion a little bit because it has an important implication for shared aspects in language and music. First of all, the unification approach is a lexicon-centered approach because all words and rules are pieces of structures (or schemas) stored in the lexicon and the unification which put those pieces of structures together is the only procedural operation (Hagoort, 2005; Jackendoff, 2002, 2011, 2015). Moreover, it also assumes a constraint-based, interactive one-stage variable-choice parsing model with phonological, syntactic, and semantic unification operating in parallel (Hagoort, 2005; Jackendoff, 2002, 2011, 2015). Finally, as already noted in the PART I, Jackendoff's parallel architecture theory, which Hagoort (2005, 2013, 2016) and Patel (2013) refer to, is understood in terms of a working memory theory concerning language processing (Jackendoff, 2002).

Concerning the DLT and the MUC, therefore, Patel's SSIRH is framed in a working memory theory although he seems to keep distance from it by saying that "[t]he approach taken here […] does not propose that […] linguistic and musical syntax share a special memory system or symbol manipulation system" (Patel, 2008, p. 276). Importantly, "working memory" is understood neither as a mere static temporary storage nor a temporary activation of elements in the long-term memory. Instead, Gibson (2000) suggested that both storage and integration components make use of the same working memory resources, and Jackendoff (2002, p. 207) emphasized that working memory is "not just as a shelf where the brain stores material, but as a workbench where processing goes on, where structures are constructed". Therefore, working memory can be regarded as a dynamic workbench or workspace where elements are temporarily maintained and integrated into larger units or sequences. In this framework, from a theoretical perspective, "shared syntactic integration resource" can be interpreted as working memory resource.

The hypothesis about shared working memory resource in language and music was articulated explicitly by Kljajevic (2010). Kljajevic (2010, p. 96) introduced the concept of syntactic working memory as "an interface module that enables processes of storage and manipulation of intermediate syntactic representations while building full representations". "Interface module" is used to indicate that syntactic working memory operates at an intermediate level between short-term and long-term memory, on one hand, and that it communicates information from different modalities and domains, on the other hand. This corresponds to Jackendoff's (2000, 2002) interface modules communicating information between different domains. "Storage" and "manipulation" correspond to storage and integration component by Gibson (1998, 2000). Kljajevic (2010), then, proposed that this syntactic working memory is also required for structural integration of musical (as well as arithmetic and action) sequences given rule-based manipulation of discrete elements. According to Kljajevic (2010), language and music syntactic representations which are built during parsing are transient, intermediate structures as sentence and musical sequence unfold in time and, thus, cannot be stored in the long-term memory, but should be represented in the (syntactic) working memory.

Beyond the theoretical parallels, there is some indirect empirical evidence for the shared working memory resource in language and music. In language research, object-extracted relative clauses and object wh-questions claim more working memory resource than subject-extracted relative clauses and subject wh-questions. In both cases, difference in working memory load is based on the distance between dependent elements. In a behavioral study

conducted by Fedorenko, Patel, Casasanto, Winawer, and Gibson (2009) which used sung melodies, comprehension accuracy of object-extracted relative clause sentences was lower than that of subject-extracted relative clause sentences and the accuracy difference between those conditions was significantly larger if the melody contained an out-of-key tone. An fMRI experiment conducted by Kunert and colleagues (2015) which used similar stimuli revealed the interaction effect between types of relative clause (subject-extracted vs. object-extracted) and pitch (in-key vs. out-of-key) in Broca's region.

10.2 Cognitive control approach

Another process involved in parsing resource intensive sentences can be investigated in terms of cognitive control, especially of "overriding highly regularized, automatic processes" (Novick, Trueswell, & Thompson-Schill, 2005, p. 263). For example, in processing one-by-one incoming words, preferred parsing strategy might point toward an incorrect analysis. Example (10-5) contains syntactic ambiguity. After reading or listening the first words "The man accepted …" people tend to expect a direct object (i.e. something accepted by the man) like "The man accepted the money." automatically. However, at the moment when the word "could" come into play, the primary parsing to take the word "the money" as a direct object turns wrong - it is actually the subject of the subordinate clause.

(10-5) The man accepted the money could not be spent yet.

In such garden-path situations, "readers and listeners have to re-rank their initial parsing commitments" (Novick et al., 2005, p. 269). The garden-path recovery involves cognitive control processes because the previous automatic and "habitual" analyses should be overridden or updated to avoid misinterpretation of the whole sequence and can be regarded as analogous to the nonparsing cognitive control tasks (e.g. stroop task[36]) including conflict resolution (Novick et al., 2005). While detection of response-based conflict was ascribed to anterior cingulate cortex, cognitive control in sentence parsing requiring detection and resolution of internal (representational) conflict was suggested to be implemented in the left inferior frontal gyrus including Broca's region (Novick et al., 2005; Novick, Trueswell, & Thompson-Schill, 2010).

[36] In the stroop task, subjects view color names presented in various ink colors and are instructed to name the color of the ink. The stimuli can be congruent (i.e., the color name and the to-be-named ink color are same) and incongruent (i.e., the color name and the to-be-named ink color are different). The task demand is larger in the incongruent condition as the participants are exposed to two conflicting representations.

In one fMRI experiment, Mestres-Missé and colleagues studied neural implementation of conflict detection and resolution in sentence parsing more in detail (Mestres-Missé, Turner, & Friederici, 2012). There are two major findings of this experiment. First, the authors predicted that processing ambiguous sentences is cognitively more demanding than processing ungrammatical sentences because the former requires both conflict detection and resolution while the latter entails only conflict detection. An evidence for this prediction was provided by behavioral results showing that the reaction time of grammaticality judgement was longer for ambiguous sentences then for ungrammatical sentences. In addition, the fMRI result showed that ambiguous sentence processing in comparison with ungrammatical sentence processing demonstrated increased activations in right superior temporal sulcus, bilateral anterior caudate head, left posterior superior temporal gyrus, and left middle temporal gyrus. Reverse contrast did not reveal any significantly more activated regions. This indicates that an additional process, i.e., conflict resolution, is required for processing ambiguous sentences. Second, this study reported two types of cognitive control gradients: a prefrontal cortex gradient ranging from dorsolateral, ventrolateral to anterior lateral prefrontal regions and a posterior-to-anterior gradient in dorsomedial striatum.

Recently, cognitive control was suggested to be one of promising candidates for resources shared in language and music processing (Slevc & Okada, 2015). Slevc and Okada (2015) point out that interaction effect (e.g., interference effect) between language and music processing which is regarded as an evidence for shared resources mainly occurs when experimental manipulations entail both conflict detection and resolution in language and music processing. For example, interaction effects are reported by behavioral and neuroimaging experiments which use musical sequences with tonal-harmonic violations and sentences with syntactic violations, non-canonical word order, or garden-path (Fedorenko et al., 2009; Fiveash & Pammer, 2014; Hoch, Poulin-Charronnat, & Tillmann, 2011; Koelsch, Gunter, et al., 2005; Perruchet & Poulin-Charronnat, 2013; Slevc et al., 2009; Steinbeis & Koelsch, 2008). Slevc and Okada (2015) further suggest that shared cognitive control resource for language and music processing is located in the left inferior frontal cortex among other cognitive processes by referring to neuroimaging studies on tonal-harmonic syntactic processing which demonstrated bilateral lateral prefrontal areas (Koelsch, Fritz, et al., 2005; Koelsch et al., 2002; Oechslin et al., 2013; Seger et al., 2013; Tillmann et al., 2003, 2006) as well as studies which showed greater grey matter density of the left inferior frontal gyrus in musicians than in nonmusicians (Abdul-Kareem, Stancak, Parkes, & Sluming, 2011; Gaser & Schlaug, 2001; Sluming et al., 2002).

Slevc and Okada (2015) claim that the most direct evidence for involvement of cognitive control resource in music processing comes from two studies which demonstrated effect of music on the stroop effect (Masataka & Perlovsky, 2013; Slevc, Reitman, & Okada, 2013). The study conducted by Masataka & Perlovsky (2013) found that parallel presentation of music consisting predominantly of dissonant intervals intensifies the stroop effect, while parallel presentation of music consisting mostly of consonant intervals does not have any effect. Slevc and colleagues (2013) also showed that parallel presentation of out-of-key endings enhanced the stroop effect, while in-key endings did not. Although those experiments use different musical stimuli, one common aspect which was hypothesized to enhance the stroop effect was chords which do not fit to the current tonal context in dissonant form (Masataka & Perlovsky, 2013) or consonant form (Slevc et al., 2013). Thus, those experiments, at least, seem to provide an evidence for the involvement of cognitive control resource in processing out-of-key chords, i.e. processing of key-membership. From those experiments, however, it does not become clear whether resource for conflict detection or resolution interact in music and stroop task.

10.3 Executive function as shared aspects of linguistic and musical syntax

Neurocognitive psycholinguistics and psychomusicology approaches take intertwined relationship between linguistic and psychological concepts. That is, they rather take a convergence strategy (see PART I, Section 5.2.1, p. 24) and explain phenomena such as grammaticality, complexity, and ambiguity by means of working memory or cognitive control. For example, complexity and ambiguity was discussed in terms of resource intensity for working memory. Conflict management in processing syntactic violation and ambiguous sentence was discussed in relation to cognitive control. The main idea concerning the relationship between language and music is that their shared aspects can be explained by means of domain-general mechanisms independent of linguistic concept such as syntax. However, by just replacing “syntax” with “working memory” or “cognitive control”, the mechanisms yielding syntactic computation cannot be revealed. Fernandez-Duque (2009), for example, suggested to investigate syntactic complexity (in particular, that of relative clauses) in terms of executive function including cognitive processes such as maintenance and manipulation of representations in the workspace, coordination of conflicting information, and switching between alternative interpretations.

For example, maintenance and manipulation of representations in the workspace corresponds to the idea of syntactic working memory engaged in both language and music. To

build a hierarchical structure, there should be a temporal storage of intermediate results (Fitch & Martins, 2014; Kljajevic, 2010; Koelsch, 2012b). Further, given conflicting hierarchical structures being represented in the workspace in parallel (Jackendoff, 2002, 2007), coordination of and switching between those conflicting representations during sentence and music parsing might also interact. The clarification of how such processes solve computational problems of language and music, e.g., processing structural relationships, will take comparative endeavors some steps forward. Especially, in musicology research, this line of investigation is still missing. Slevc and Okada (2015) listed up which musical phenomena parallel garden-path effect of language. This is an important first step toward a theory of music syntactic processing.

In addition, the argument from cognitive neuroscience perspectives is often made by a reverse inference as follows: if executive function is engaged, the left inferior frontal gyrus is activated; language and music processing activate inferior frontal gyrus; that is, a shared aspect of language and music processing is executive function. Again, the selectivity of the left inferior frontal gyrus for executive function is not strong enough to make this reverse inference. Moreover, a fine-grained analysis of the inferior frontal gyrus reveals non-overlap between language and music syntactic regions and executive regions. For example, within Broca's region, the most dorsal portion of BA 44 is strongly associated with working memory (Clos et al., 2013). This is in line with a study dissociating storage component of working memory and nested dependencies (Makuuchi et al., 2009). The area within BA 44 which is associated with cognitive control was in the region of the inferior frontal junction (Clos et al., 2013).

11 Shared mechanisms for syntax in language and music

Experiments which provided evidence for the hypotheses about shared neural resources for syntacitic processing in language and music (e.g., Koelsch, Gunter, et al., 2005; Kunert et al., 2015; Sammler et al., 2009; Slevc et al., 2009; Steinbeis & Koelsch, 2008) mainly tested interaction and neural overlap as implications on the basis of following inferences: 1) Sharing → Interaction; 2) Sharing → Overlap. The results of the review reveal a paradox. On one hand, interaction studies show that language and music syntactic processing rely on cognitive processes involved in recognition of structural relationship and executive function. On the other hand, evidences for neural overlap are very weak. For example, while some studies (Kunert et al., 2015; Sammler et al., 2009) provide evidence for overlap between language and music syntactic processing, the meta-analysis conducted in the current thesis shows that regions (consistently) associated with music syntactic processing little overlap with 'language

regions'.[37] Especially, the overlap disappears if the region (e.g., Broca's region) is subdivided into smaller areas. Therefore, the basic premise of current comparative language-music research "Sharing → Overlap" showed its limitation.

Patel (2013, p. 336) introduced three possible hidden connections between language and music processing in the brain to account for such a paradox:

1. The same brain network, but developmentally one domain (language, music) is much more robust to impairments in this network.
2. The interaction of shared brain networks with distinct, domain-specific brain networks.
3. Separate but anatomically homologous brain networks in opposite cerebral hemispheres.

Patel (2013, p. 336) suggested that "musical and linguistic cortical processing rely on a similar functional computation, yet musical and linguistic abilities can be dissociated by brain damage" in each of those three cases. The first case, for example, explains why subjects with congenital amusia show more pronounced deficits in musical pitch processing than in speech, although speech and music rely on the same brain networks for relative pitch processing. The second case is the core idea of SSIRH, which explains the relationship between linguistic syntactic and tonal-harmonic syntactic processing in terms of shared resource networks, but different representation networks. The third case was suggested to potentially explain the relationship between speech and song motor control in the brain.

The result of the meta-analysis on tonal-harmonic syntactic processing suggests the third hidden connection between syntax in language and music. Tonal-harmonic syntactic processing recruits a right fronto-temporal network including the right inferior frontal gyrus (BA 44) and the right posterior superior/middle temporal gyrus (BA 22). In addition, current comparative language-music research on syntax largely focuses on cortical structures and neglects hidden connections in subcortical structures. For example, Seger and colleagues (2013) showed that the basal ganglia's activities were significantly modulated by harmonic violations in music. Moreover, the neural generator of P600 was suggested to include the basal ganglia (Friederici & Kotz, 2003). Therefore, comparative language-music research could

[37] It is worth noting that Kunert and colleagues (2015) as well as Sammler and colleagues (2009) use within-subject design, while the meta-analysis is conducted by taking multiple subjects from differently designed experiments. Thus, this is a basic methodological issue which requires further considerations if we discuss neural overlaps.

benefit from taking the role of the subcortical structures into account (see, for example, Kotz, Schwartze, & Schmidt-Kassow, 2009).

Another strategy to tackle the paradox is giving up the conceptual dichotomy of sharing and non-sharing. Instead, language and music can be best investigated in terms of a continuum (Arbib, 2013). For example, within the left BA 44, there could be a dorsal-ventral axis from working memory in the most dorsal part via the part associated with language to the most ventral 'action and music' part. However, the function such as maintenance and manipulation could be same for all of them. Similarly, there is a rostro-caudal gradient of memory, control, and goal representation in the frontal cortex with motor part in the most caudal part and cognitive or abstract part in the most rostral part (Badre & D'Esposito, 2009; Badre & Nee, 2018; Fuster, 2008b; Koechlin & Jubault, 2006; Rouault & Koechlin, 2018; Uddén & Bahlmann, 2012). Language and music can be also differently represented on this axis. For example, music syntactic processing rarely extends to BA 45 in contrast to language syntactic processing which rely on BA 45 consistently.

Moreover, generative neurolinguistics and neuromusicology as well as neurocognitive psycholinguistics and psychomusicology suggest hypothetical components shared in language and music syntactic processing such as hierarchical processing or executive function, but they don't explicitly state how the mind/brain solves the computational problem. That is, they do not provide hypotheses about cognitive and neural processes and do not approach mechanistic questions. On one hand, computational-representational theories cannot be ignored because they provide explanations to given phenomena. However, a purely top-down approach has its limit if the target of research are neurocognitive mechanisms. What are cognitive processes which can account for the computational problem to be solved and can be implemented by neural processes at the same time? A complementary bottom-up approach could take us way forward. For example, one could take the brain regions which are hypothesized to be involved in language and music syntactic processing, and study on the basis of which implementational principles information is processed in those regions.

For example, the prefrontal cortex, together with the basal ganglia, carry out inhibitory control (Aron, Robbins, & Poldrack, 2014; Fuster, 2008b; E. K. Miller & Cohen, 2001) to alter on-going action or representation. In the prefrontal cortex, delay neurons provide prerequisite to maintain information (Fuster, 2008b; E. K. Miller & Buschman, 2008) to process relationship between elements. Thus, control and maintenance are cognitive processes which can be implemented by the brain and can also account for some aspects of hierarchical

processing and executive function hypothesized to be shared in language and music syntactic processing.

Part III Rhythmic syntax: An integrative approach

Part III extends the discussion on the relationship between syntax in language and music by introducing the concept of rhythmic syntax, which is an often-neglected aspect of musical syntax. In the first chapter, I first argue that musical rhythm is best investigated in terms of syntax. I further identify components of musical syntax and elucidate how those components work together to link sound and affect. Finally, I conclude by proposing operating principles of rhythmic syntax. The second chapter deals with rhythmic syntactic processing and its neural correlates. In addition, I introduce a hypothesis about neurocognitive mechanisms for rhythmic syntactic processing. In the third chapter, then, the relationship between language and music is discussed in light of rhythmic syntax, rhythmic syntactic processing, and the underlying neurocognitive mechanisms. The final, fourth chapter deals with the relationship between syntax in language and rhythmic syntax.

12 Computational-representational theory of rhythmic syntax

12.1 Components of rhythmic syntax

In current comparative language-music research, as reviewed and discussed in PART II, linguistic syntax has been compared with tonal-harmonic syntax in music (e.g., Koelsch, 2011a, 2012a; Patel, 2003, 2008, 2012, 2013; Rohrmeier, 2011; Tillmann, 2012). However, links from the musical domain to linguistic syntax are not limited to tonal-harmonic syntax. Musical rhythm is an organizing principle of music (Asano & Boeckx, 2015; Fitch, 2013; Longuet-Higgins & Lee, 1984; Patel, 2008) and thus is an integrative part of musical syntax. As the central computational problem of music is linking sound and affect, rhythmic syntax then should account for it solely in terms of temporal organization. In the following, I discuss which components constitute rhythmic syntax as well as what they do and how they work together to link sound and affect. Musical rhythm entails two distinct, but interacting subsystems, namely grouping and meter (Lerdahl & Jackendoff, 1983). Importantly, both subsystems organize auditory events hierarchically. Interaction between those two subsystems results in an asymmetrical headed hierarchy representing structural importance and affect.

Grouping refers to a hierarchical organization of the musical stream into motives, phrases, and sections (Lerdahl & Jackendoff, 1983). The boundaries of groups at the phrase level and all larger grouping levels were suggested to mark structural beginning and ending (or cadence) called “structural accents”. Thus, it can be seen as segmentation of the musical stream into units. A group is the basic unit of grouping and can be constituted by any contiguous auditory event. Hierarchical grouping structures can be called ‘recursive’ in a way that is similar to visual grouping (Jackendoff, 2009). In Figure 12.1, a set of ‘x’ forms a line, a set of lines forms a small rectangle, and a set of small rectangle forms a larger rectangle. Jackendoff and Pinker (2005) suggested that this kind of visual grouping is recursively and hierarchically organized. Such groupings are recursive in the sense that the results of lower hierarchical levels are the inputs to higher hierarchical levels, i.e., a form of self-reference. The groups of the lower levels are contained in the groups of the higher levels, which means that the lower-level groups are subordinate to the higher-level groups and the higher-level groups dominate or are superordinate to the lower-level groups. Thus, the result of grouping is a layered hierarchical structure with multiple hierarchical levels that are in a subordinate-dominate relationship.

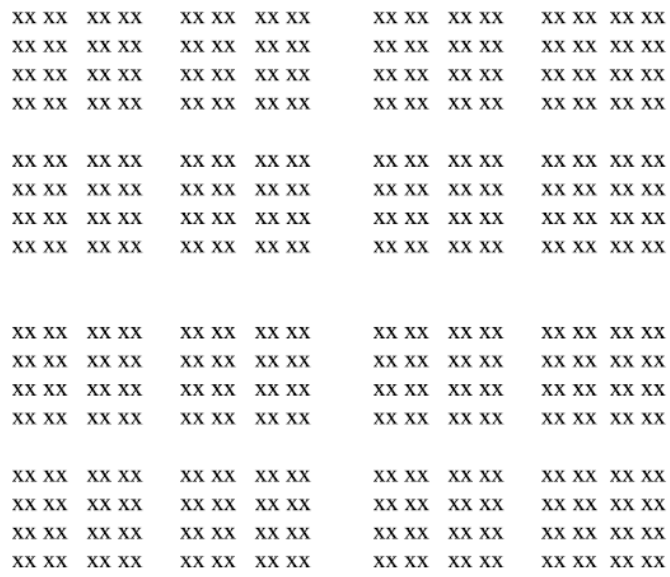


Figure 12.1 Visual grouping. Adopted from Jackendoff & Pinker (2005, p. 218), *Cognition*, 97: 211-225, with permission by Elsevier.

Meter generates a regular pattern of strong and weak beats called "metrical accent" that is represented as hierarchical metrical structure (Lerdahl & Jackendoff, 1983). The basic elements of meter are beats, which, "as such, do not have duration" (Lerdahl & Jackendoff, 1983, p. 18). Thus, beats are analogous to geometric points rather than to lines. Importantly, metrical accent and beats are mental constructs that are inferred from, but not identical to the patterns of phenomenal accent, i.e., accentuation at the musical surface.[38] In metrical structures, beats are organized hierarchically according to their relative strength (see *Figure 12.2*). Hierarchical organization of metrical structure represents strong and weak beats at different levels. As shown in *Figure 12.2*, metrical structure is often notated in a metrical grid, in which strong beats project onto higher levels. In (western) music, the beats are equally spaced, i.e., isochronously organized, in a sequence and the pattern of strong and weak beats, i.e., metrical accent, is periodic (Lerdahl & Jackendoff, 1983; Patel, 2008). In binary meter, two beats are assembled at each level and in ternary meter three beats. Metrical structure is recursively organized in the sense that the results of lower hierarchical levels are the inputs to higher hierarchical levels, which is, again, a form of self-reference.

[38] Accentuation at the musical surface includes, for example, "attack points of pitch-events [or auditory events; RA], local stresses such as sforzandi, sudden changes in dynamics or timbre, long notes, leaps to relatively high or low notes, harmonic changes, and so forth" (Lerdahl & Jackendoff, 1983, p. 17). In the current thesis, in most of the case, "phenomenal accent" refers to attack points of auditory events.

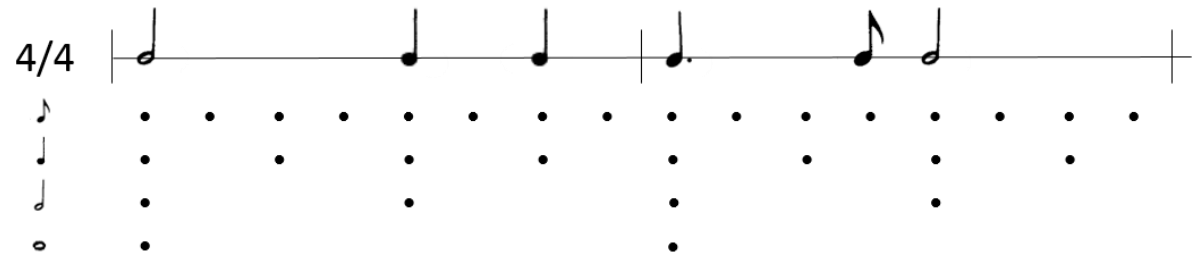


Figure 12.2 Metrical grid representing beats at multiple hierarchical levels.

The hierarchical metrical structure is different from that of grouping: metrical structure is organized asymmetrically in form of (quasi-)headed hierarchy (Fitch, 2013; Jackendoff, 2009; Lerdahl & Jackendoff, 1983). That is, metrically stable and important elements, i.e. heads, serve as reference points of hierarchical organization. The heads are indicated as higher projecting beats in *Figure 12.2*. In music, such a metrical hierarchy contains a particular hierarchical level called "tactus" (Lerdahl & Jackendoff, 1983, p. 21). Tactus is the level where listeners tap their foot or clap their hands. It shows the most stringent regularities of metrical structures. This is also the level of the shared beats for coordinating ensemble and dance performance. That is, listeners usually focus on the beats at one level of hierarchical metrical structure, i.e., the tactus, and do not perceive the beats at all hierarchical levels in parallel. Below, I call the beats at the tactus level "primary beats".

Hierarchical structure of musical rhythm, then, is determined on the basis of both components, i.e., meter and grouping, in form of an asymmetrical headed hierarchy representing structural importance called "time-span structure" (Lerdahl & Jackendoff, 1983, p. 146).[39] Musical sequence is segmented into hierarchical layers ranging from micro to macro time units. This is what is called "time-span segmentation" in GTTM (Lerdahl & Jackendoff, 1983). A time-span is a rhythmic unit and an interval of time between successive beats at the smallest level. At the smallest level, time-spans are built solely on the basis of metrical structure and are regular in length. At the intermediate levels, regularity of time-spans imposed by metrical structure is interrupted by grouping boundaries. At larger levels, then, grouping

[39] "Time-span structure" is defined as "the segmentation of a piece into rhythmic units within which relative structural importance of pitch-events can be determined" (Lerdahl & Jackendoff, 1983, p. 146). "Time-span reduction" is another term which is possibly better known in the context of *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983 Chapter 6). The concept of "reduction" strongly implies "a step-by-step simplification […] of a piece" (Lerdahl & Jackendoff, 1983, p. 106). Because I do not intend to emphasize the step-by-step procedure in the current thesis in developing a computational-representational theory, I use "time-span structure" which rather points out hierarchically structured representation. In the Figure 12.3, the idea of the reduction is represented below as (a) → (b) → (c) and so on. The same applies to "prolongational structure" and "prolongational reduction".

completely takes over determining role. In this way, time-spans are organized hierarchically from small to large level (Figure 12.3 bottom). Within each time-span at each hierarchical layer, a structurally most important event is chosen as a head. In Western tonal music, the heads are selected on the basis of metrical strength, group boundary, and melodic and harmonic stability. Such a time-span structure is represented at the top of Figure 12.3. That is, meter and grouping not only segment the musical stream into rhythmic units, i.e., time-spans, but also determine the structural importance of auditory events.

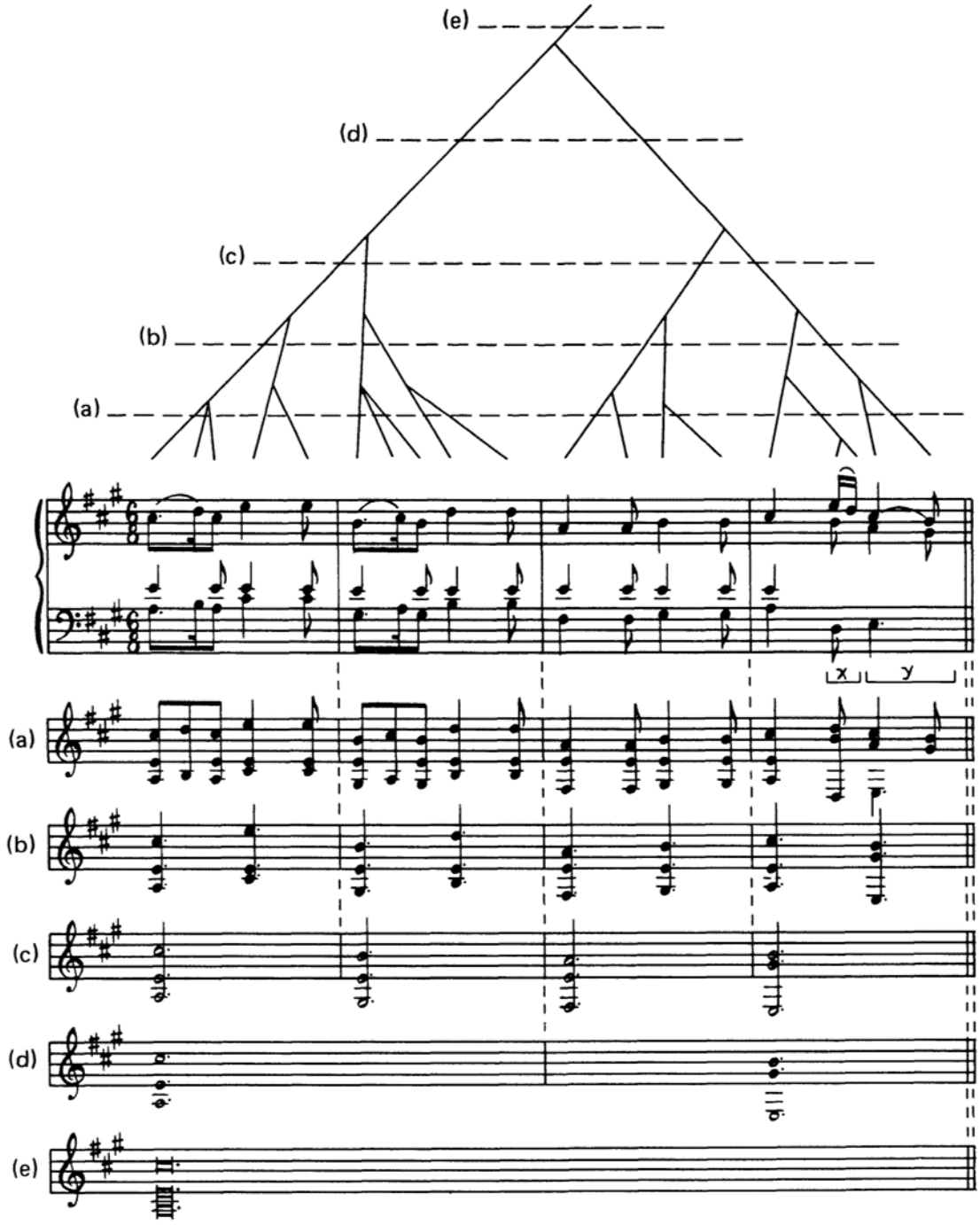


Figure 12.3 Time-span structure. The figure is adopted from Jackendoff (1987, p. 226) with permission: Credit to Ray S. Jackendoff, *Consciousness and the Computational Mind*, reprinted courtesy of The MIT Press.

It is reasonable to consider melodic and harmonic stability in the time-span structure of tonal music because tonal-harmonic structure is central to it. However, this stability criterion cannot be applied to rhythmic music such as Western African drumming music. To build up a time-span structure of music which does not rely on pitch stability condition, we therefore need other criteria. In addition, time-span structure does not encode affect and prolongational structure encoding tension-relaxation pattern centers on tonal motion. Tension-relaxation patterns that are built solely by rhythm were neglected in GTTM. Because the link between sound and affect is the central computational problem to be solved in music, any theory of rhythmic syntax should account for building up tension-relaxation patterns in musical pieces which do not rely on pitch stability conditions (see PART II, Section 8.2.1, p. 51).

Lerdahl (1989) introduced salience conditions to account for time-span and prolongational structure independent of pitch stability conditions. Salience conditions for determining heads in time-span structure includes local conditions such as 1) corresponding to attack, i.e., sound onset, 2) at metrically strong position, 3) relatively loud, 4) relatively prominent timber, 5) in an extreme registral position (high or low), 6) relatively dense, and 7) relatively long in duration as well as global conditions such as 8) motivically relatively important, 9) grouping boundary, and 10) parallel to a choice elsewhere in the analysis. The local conditions can be explained in terms of metrical and phenomenal accents. The global conditions largely relate to grouping and structural accents.

Based on the salience conditions, I suggest that an auditory event is maximally stable if phenomenal, metrical, and structural accents correspond to each other and maximally instable if there is no correspondence between them. Instable auditory events cause tension. There are several phenomena in musical rhythm that are known to cause tension based on salience conditions: syncopation and polyrhythm (Fitch, 2016; Vuust, Gebauer, & Witek, 2014; Vuust & Witek, 2014). For example, syncopation occurs when phenomenal and metrical accents do not match. Thus, from a computational perspective on rhythmic syntax, music that makes use of those phenomena to create tension-relaxation pattern is of particular interest. One example of such music is West African drum ensemble music. As we will see below, tension-relaxation patterns in West African drum ensemble music can be analyzed by applying such

salience conditions in terms of the relationship between structural, metrical, and phenomenal accents.

12.2 Structural analysis of African drum ensemble music

West African drum ensemble music emphasizes the role of timeline patterns (also called "standard patterns") which is usually played on instruments with timbres distinct from the rest of an ensemble (Polak, 2010). The pattern shown in Figure 12.4 is a timeline pattern of Ewe music called (slow) Agbekor in Ghana. The lowest level of metrical structure is inserted at the eighth note level so that all attacks on the musical surface are associated with a beat. This lowest subdivision level of meter containing 12 beats is the baseline for every rhythmic activity (Locke, 1982). Then, strong beats are projected onto the higher level which is the tactus of this standard pattern.[40] This interpretation is represented in the foot movements of West African dance (Agawu, 2006; London, 2012a) and also played by rattles which have very prominent timbre. Most authors seem to agree on the existence of such a particular beat cycle underlying polyrhythm (Polak, 2010). Those four beats were claimed to receive equal metrical accents (Locke, 1982). However, at least one higher metrical level called "metric cycle" or "metrical cycle" can be assumed because the first beat of the timeline pattern is considered as the downbeat (Polak, 2017; Temperley, 2000). There is some very weak evidence for an additional level between tactus and metrical cycle (Temperley, 2000). Nevertheless, this additional level is plausible as African performers clap at this level, too.

Therefore, there are at least three hierarchical levels of metrical structure (and very likely also a fourth level) as displayed in Figure 12.4 in metrical grid notation. The primary beats are generated on the basis of ternary organization of beats at the lowest subdivision level. The metrical level above the tactus is then based on binary organization of the primary beats. Finally, the metrical cycle level organization is also based on binary organization. That is, to construct the metrical hierarchy, beats are first combined into ternary units and then into binary units.

[40] The metrical ambiguity which is inherent in this standard pattern as well as the possibility of non-isochronous meter are discussed in Section 13.2.

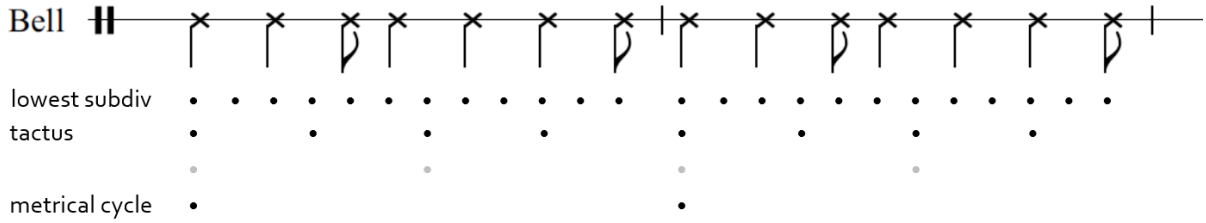


Figure 12.4 Metrical structure underlying the timeline pattern of Agbekor. The level between tactus and metrical cycle is in gray because there is no consensus among researchers whether this structural level is relevant or not.

This metrical structure is well-formed according to a set of rules suggested by Lerdahl and Jackendoff (1983) to generate metrically well-formed structures.

(12-1) Metrical well-formedness rules (MWFR) (Lerdahl & Jackendoff, 1983, p. 347)

MWFR 1 (**revised**) (p. 72)
Every attack point must be associated with a beat at the smallest metrical level present at that point in the piece.

MWFR 2 (**revised**) (p. 72)
Every beat at a given level must also be a beat at all smaller levels present at that point in the piece.

MWFR 3* (p. 68)
At each metrical level, strong beats are spaced either two or three beats apart.

MWFR 4 (**revised**)***** (p. 72)
The tactus and immediately larger metrical levels must consist of beats equally spaced throughout the piece. At subtactus metrical levels, weak beats must be equally spaced between the surrounding strong beats.

First, as every attack point must be associated with a beat at the smallest metrical level according to MWFR 1, the lowest subdivision level is at the eighth note level as shown in Figure 12.4. Second, according to MWFR 2 and 3, every two or three beat at a given level (*l*)

can be projected onto the higher level (l + 1). In Figure 12.4, every three beat at the lowest subdivision level projects onto the tactus level and every two beat at the tactus level projects onto the higher level (*tactus* + 1), corresponding to what MWFR 2 and 3 state. In accordance with MWFR 4, beats are equally spaced at the tactus and immediately larger metrical levels.

Some readers might have recognized that MWFR 2 and 3 lead not only to metrical structure represented in Figure 12.4, but also to other possibilities. I will elaborate this issue later in Section 13.2 (p. 122). For now, I only discuss whether this particular metrical structure in Figure 12.4 can be supported by metrical preference rules (MPR) introduced by Lerdahl and Jackendoff (1983). Among a series of MPRs, Lerdahl and Jackendoff (Lerdahl & Jackendoff, 1983, pp. 278–279) suggested MPR 4 (Stress) as applicable to a wide range of musical idioms. MPR 4 (Stress) states "[p]refer a metrical structure in which beats of level L_i that are stressed are strong beats of L_i" (Lerdahl & Jackendoff, 1983, p. 79). What does "stressed" mean in African music? As Agbekor is a dance music, one possibility is regarding dance movements as placing stresses to beats of the timeline pattern. This interpretation is in line with the claim that the primary beats represented in the foot movements of West African dance (Agawu, 2006; London, 2012a). Every three beat at the lowest subdivision level projects onto the tactus level because dance movements put stress on every three beat.[41] In addition, the first beat of the timeline pattern is often stressed by large dance movement and lower pitched bell sound, leading to the metrical representation where the first beat of the timeline pattern gets the strongest metrical accent. Of course, there are other factors which put stress on the timeline pattern to prefer other metrical representations, which I will discuss in Section 13.2 (p. 122).

A grouping structure which can be represented without controversies (for discussions, see Agawu, 2006) is the whole pattern, which is repeated during the entire performance. Grouping structure of West African drum ensemble music is often notated in form of the number of time-spans at the lowest subdivision level which fall into one inter-onset interval (IOI) (Figure 12.5). This is also called "interval-vector" (Pressing, 1983). For example, the first IOI of the timeline pattern of Agbekor contains two time-spans and notated as 2, while the third IOI contains only one time-span and is thus notated as 1. Concerning the subgroups, the timeline pattern seems to additively consist of a pattern of 5 beats organized in 2+2+1 and a pattern of 7 beats in 2+2+2+1. In this way, two or three long strokes are separated by one short stroke (Agawu, 2006; Locke, 1982). Although 5+7 organization is one possible interpretation

[41] Agbekor dance performance consists of multiple dance figures. Only some of the figures is in support of this view. See also the discussions concerning the relationship between grouping structure and dance figure below.

of the timeline pattern on the basis of the analysis of IOIs at the musical surface, 7+5 organization is more plausible (Figure 12.5 top). For example, the timeline pattern can be also realized in five stroke version in form of 2+2+3+2+3 (Agawu, 2006) indicating 7+5 organization rather than 5+7. Moreover, 2+2+3+2+3 patterns are more common in African music than 2+3+2+2+3 patterns (London, 2012a). In addition, in contrast to grouping preference in Western music, African music favors strong beats occurring at the ends of groups (Temperley, 2000), which is true for 7+5 grouping.

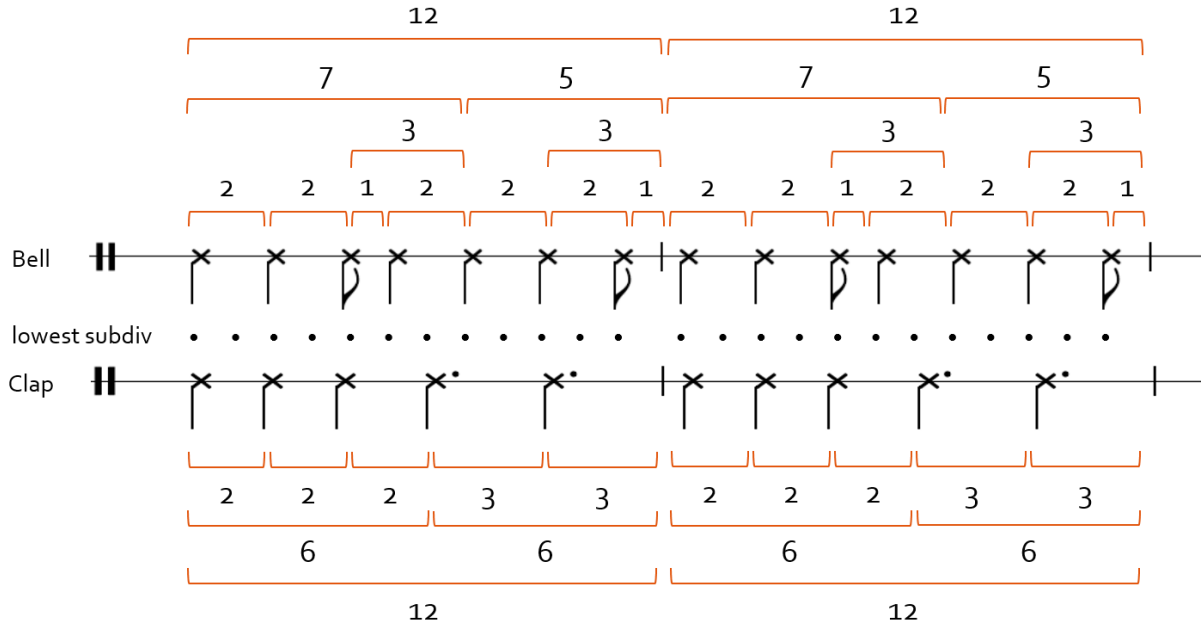


Figure 12.5 Grouping structures underlying the timeline pattern of Agbekor.

This grouping structure (Figure 12.5 top) is well-formed according to a set of rules suggested by Lerdahl and Jackendoff (1983) to generate well-formed grouping structures.

(12-2) Grouping well-formedness rules (GWFR) (Lerdahl & Jackendoff, 1983, p. 345)

GWFR 1 (p. 37)
Any contiguous sequence of pitch-events, drum beats, or the like can constitute a group, and only contiguous sequences can constitute a group.

GWFR 2 (p. 38)
A piece constitutes a group.

GWFR 3 (p. 38)

A group may contain smaller groups.

GWFR 4 (p. 38)

If a group G_1 contains part of a group G_2, it must contain all of G_2.

GWFR 5 (p. 38)

If a group G_1 contains a smaller group G_2, then G_1 must be exhaustively partitioned into smaller groups.

GWFR 1 is a kind of prerequisite for grouping analysis and allows that all adjacent auditory events can be grouped together. GWFR 2 refers to a piece which I regard as equating the entire Agbekor performance in our example. GWFR 3 and 4 correspond to hierarchical organization of Figure 12.5 in which a group contains smaller groups and the smaller groups are entirely subsumed in the larger group. Is it possible to identify the 7+5 organization of grouping structure according to grouping preference rules (GPR)?[42] First of all, the timeline pattern can be regarded as a group on the basis of GPR 6 (Parallelism): "Where two or more segments of the music can be construed as parallel, they preferably form parallel parts of groups" (Lerdahl & Jackendoff, 1983, p. 51).

Concerning the smaller groups, a 1+2 group (i.e., a group consisting of an eighth note and a subsequent quarter note) can be formed according to the GPR 2 (Proximity) b. (Attack-Point) (Lerdahl & Jackendoff, 1983, p. 45). The interval between the attack of the eighth note (a) and the attack of the quarter note (a + 1) is smaller than the surrounding intervals between attacks, i.e., attack between the eight note (a) and the previous quarter note (a - 1) as well as that between the quarter note following the eight note (a + 1) and the subsequent quarter note (a + 2). Although the same can be said about the interval between the last eight note of the timeline pattern and the first quarter note of the next timeline pattern, there are good reasons to not regard them as a group. First, the first quarter note of the timeline pattern is often played by a lower pitched bell so that a group boundary can be assigned before this quarter note according to GPR 3 (Change) a. (Register) (Lerdahl & Jackendoff, 1983, p. 46). Moreover, a

[42] Tmeperley (2000) also discusses this question. His strategy was to take the phrasing slurs used in transcriptions (Temperley, 2000, p. 89). However, this approach faces a serious problem that the slurs used in transcriptions often reflect interpretations of Western researcher and might not correspond to groupings of African music. Therefore, I suggest an alternative strategy to take dance segments to identify grouping boundaries.

new dance figure begins with the first quarter note of the timeline pattern, also indicating a group boundary before the first quarter note of the timeline pattern.

Finally, concerning the 7+5 grouping, again, dance figure boundaries should be taken into account because 'musical surface' otherwise cannot grouped unambiguously into 7+5 grouping instead of 5+7 grouping. Some segments of Agbekor dance performance contain a figure combination consisting of a large jump and three small steps (turning knee back and forth) where the jump corresponds to 7 interval vectors (i.e., 2+2+3) and three small steps correspond to 5 interval vectors (i.e., 2+2+1). This grouping by dance can be established according to GPR 3 (Change) b. (Dynamics) (Lerdahl & Jackendoff, 1983, p. 46) as the large jump figure and the small three-step figure differ in their dynamics.

Concerning grouping, in addition, there are further patterns in polyrhythmic texture of Agbekor. One very common grouping is shown by the clapping pattern: 2+2+2+3+3 (Locke, 1982; Temperley, 2000). An accompanying song melody is also grouped into this pattern according to GPR 3 (Change) a. (Register) (Lerdahl & Jackendoff, 1983, p. 46): all group boundaries in the song melody are marked by pitch changes. The 2:3 ratio of grouping is called "hemiola" (London, 2012a): the first half of the 12 beats is grouped in 2 and the latter half in 3 (Figure 12.5 bottom). This 2:3 ratio of grouping in clapping and song melody is in line with the heart of Ewe drum music, namely "each instrument relates its pattern both to the bell pattern and the primary metric accents" (Locke, 1982, p. 222). The clap pattern corresponds to the bell pattern in the first half and to the metrical accent in the second half. Although this 2+2+2+3+3 pattern is often discussed in terms of non-isochronous meter (e.g., London, 2012a), I regard it as grouping because it has to do with the IOIs, i.e., groups of time-spans, and keep the discussion about non-isochronous meter for Section 13.2 (p. 122).

The results of the above mentioned grouping analyses are shown in Figure 12.5. Figure 12.5 reveals two concurrent grouping structures, which persist in parallel to further rhythmic patterns with own grouping structure. Existence of multiple concurrent grouping structures is characteristic of polyrhythmic music. An analysis of Agbekor dance which I used to identify grouping structure can be found in Appendix B (p. 237).

From the structural analysis of meter and grouping in Agbekor drum music conducted so far, one combinatorial principle becomes apparent: both metrical and grouping structures are constructed by combining two or three elements. For example, beats are combined hierarchically in a 3*2*2 manner to build metrical structure represented in Figure 12.4. Although it is not clear whether inherent 'twoness' or 'threeness' are perceived quantitatively, they are suggested to be categorically different in their quality being short (S) or long (L)

(London, 2012a). Especially, this short-long qualitative difference is evident in the binary and ternary combination of units in a sequential order as seen in 2+2+3+2+3 or 2+2+2+3+3 patterns which can be re-written as S-S-L-S-L and S-S-S-L-L. Alternatively, Lerdahl and Jackendoff (1983), following Singer (1974), use quick (Q) and slow (S) to characterize the 2+2+3 pattern in Macedonian dance as Q+Q+S. That is, two and three can be considered as labels of combinatorial units regardless of meter or grouping.

Moreover, it is important to note that binary and ternary combinatorics is not an 'exotic' phenomenon. For example, all pieces in Western music with the time signature 6/8 make use of 3*2 or 2*3 combinatorics to build metrical structure and 3+3+2+2+2 patterns are also used in *America* of the *West Side Story* composed by Leonard Bernstein. The score in Figure 12.6 represents the 3+3+2+2+2 pattern in *America* on the basis of eight notes. In the first bar, three eight notes are tied together as a unit twice. This corresponds to a 3+3 pattern. The second bar contains three quarter notes. Because one quarter note consists of two eight notes, the pattern in the second bar corresponds to 2+2+2. Both together, then, form a 3+3+2+2+2 pattern.

In addition, the binary and ternary combinatorics is not limited to create one particular level of representation. This is used at all hierarchical levels. For example, in the African timeline pattern (Figure 12.5 top), 2+2+3+2+3 are again hierarchically organized by combining the first three units as a larger unit and the following two units as another larger unit. In addition, by regarding three as combination of two units, i.e. two and one, the surface representation of the seven-stroke timeline pattern, i.e., 2+2+1+2+2+2+1, can be derived at the surface level.

In the *West Side Story* example, higher level binary and ternary combinatorics is reflected in conductor's arm movements: 3+3+2+2+2 pattern is organized hierarchically by combining 3+3 as a binary unit and 2+2+2 as a ternary unit. Figure 12.6 (bottom) shows that the conductor shows two movement units in the first bar (Figure 12.6 bottom left) and three movement units in the second bar (Figure 12.6 bottom right). The first two movement units correspond to 3+3 and the following three movement units match 2+2+2. That is, in the *West Side Story* example, eighth notes are first combined into ternary and binary units which are again combined into binary and ternary units: $_2$[$_3$[x, x, x], $_3$[x, x, x]], $_3$[$_2$[x, x], $_2$[x, x], $_2$[x, x]].[43] Thus, the binary and ternary combinatorics is one way to characterize hierarchical structure of

[43] Elements (x) are combined into units as indicated with "[ ]", which are combined into further units. Here, "x" stands for a beat at the eighth note level. "2" stands for binary unit and 3 for ternary unit. The units indicated with "[ ]" are order sensitive, i.e., the sequential order of elements / units within "[ ]" should be as they notated. This differs from units indicated with "{ }", which does not determine the sequential order of elements / units. That is, [x, y] and [y, x] differ, while {x, y} and {y, x} are the same. To put it other words, "[ ]" notation represents both precedence and dominance relationship between elements / units, while "{ }" notation only determines dominance relationship between elements / units.

musical rhythm, which is applicable to wide range of musical culture. Further, ‘twoness’ and ‘threeness’ of units can be regarded as categories or ‘labels’ of those units.

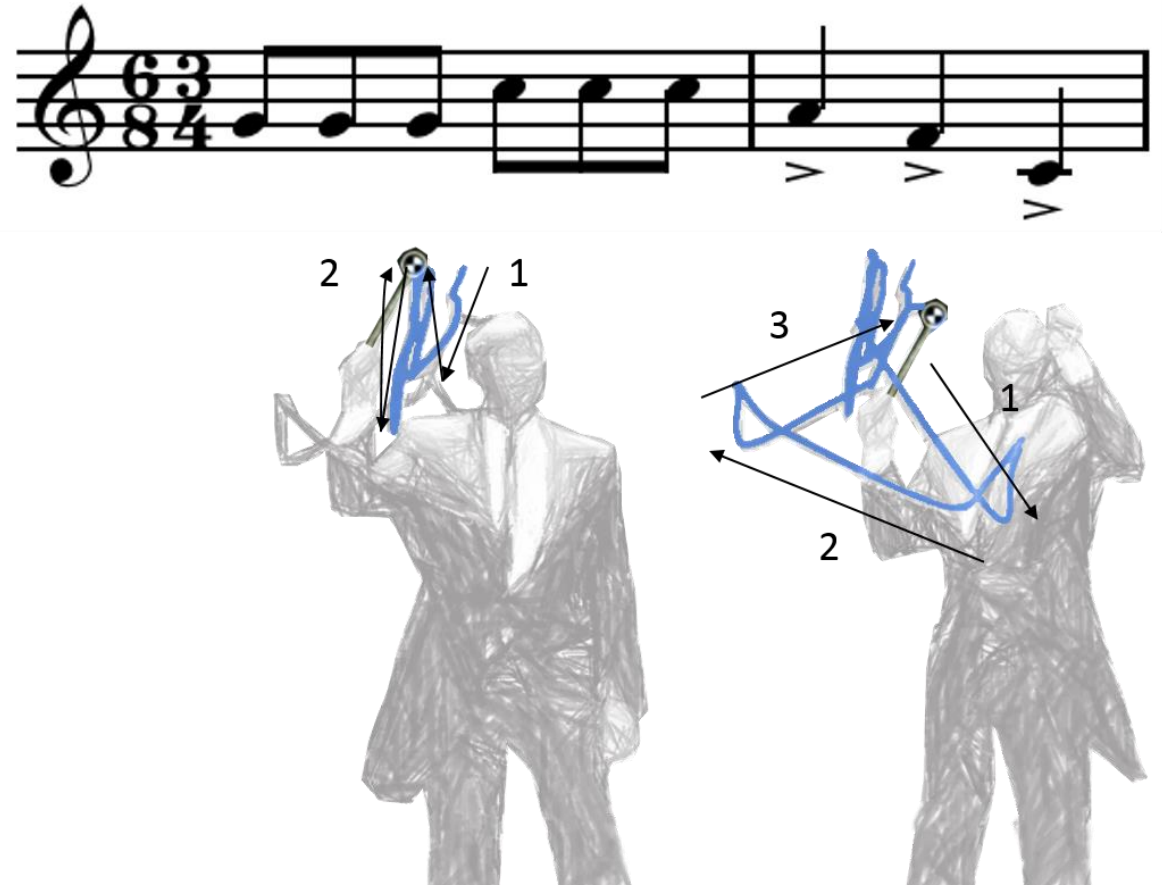


Figure 12.6 Conductor’s movement trajectory in *America* from the *Westside Story* composed by Leonard Bernstein. A part of the score is depicted at the top of the figure and the corresponding movement trajectory is represented at the bottom. The trajectory was created with Kinovea software (https://www.kinovea.org/). The marker was put on the top of the baton as indicated with the black-white circle for automatic trajectory tracking (with some manual corrections). The video (https://www.youtube.com/watch?v=_LDABUJAS1w) was retrieved from YouTube on November 1st, 2018, and converted by using RealPlayer (https://www.real.com/de). The picture was edited with GIMP software (https://www.gimp.org/).

Metrical and grouping structures of West African drum ensemble are hierarchically organized by means of binary and ternary combinatorics as discussed above. The question remains how tension-relaxation patterns are encoded on the basis of those hierarchical structures. While tension-relaxation patterns of Western tonal music are encoded on the basis

of tonal motion, some authors suggested that tension in West African drum ensemble music can be created in terms of mismatch or conflict between metrical and phenomenal accents. For example, Locke (1982) pointed out that Ewe music makes use of off-beat timing (also called "syncopation"), i.e., non-correspondence of attacks to metrical accent, to create tension, and suggested that a sense of arrival, i.e., relaxation, is created if attacks coincide with metrical accent. Temperley (2000) also notes that conflict between surface events and meter gives rise to tension. In Figure 12.7, tension-relaxation patterns of the standard pattern are encoded on the basis of the relationship between metrical accent and attacks on the musical surface, with the correspondence as stability causing relaxation and incongruity as instability causing tension. The tree notation is adopted from prolongational structure (Lerdahl & Jackendoff, 1983) and encodes different tension-relaxation patterns. The branching considers metrical structure and its relation to phenomenal accent only.

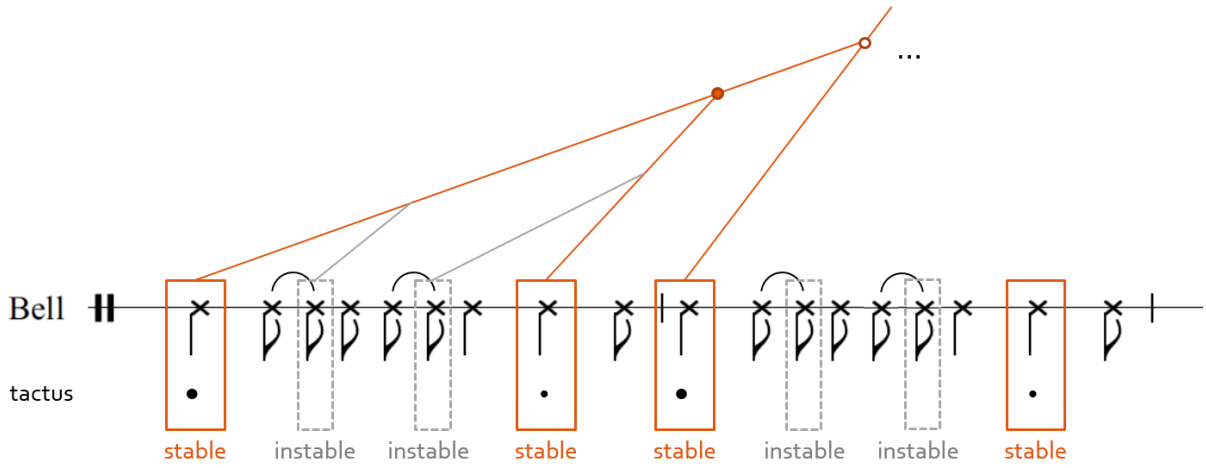
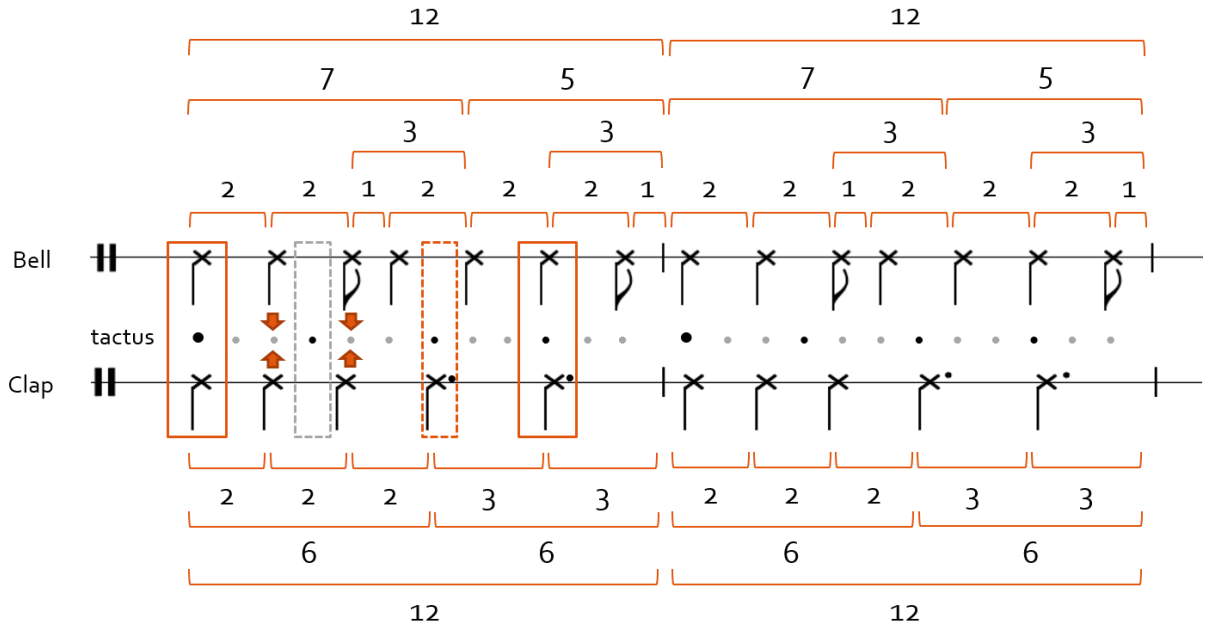


Figure 12.7 Tension-relaxation pattern of the standard pattern of Agbekor (preliminary). In this prolongational structure, the branching considers metrical structure and its relation to phenomenal accent only.

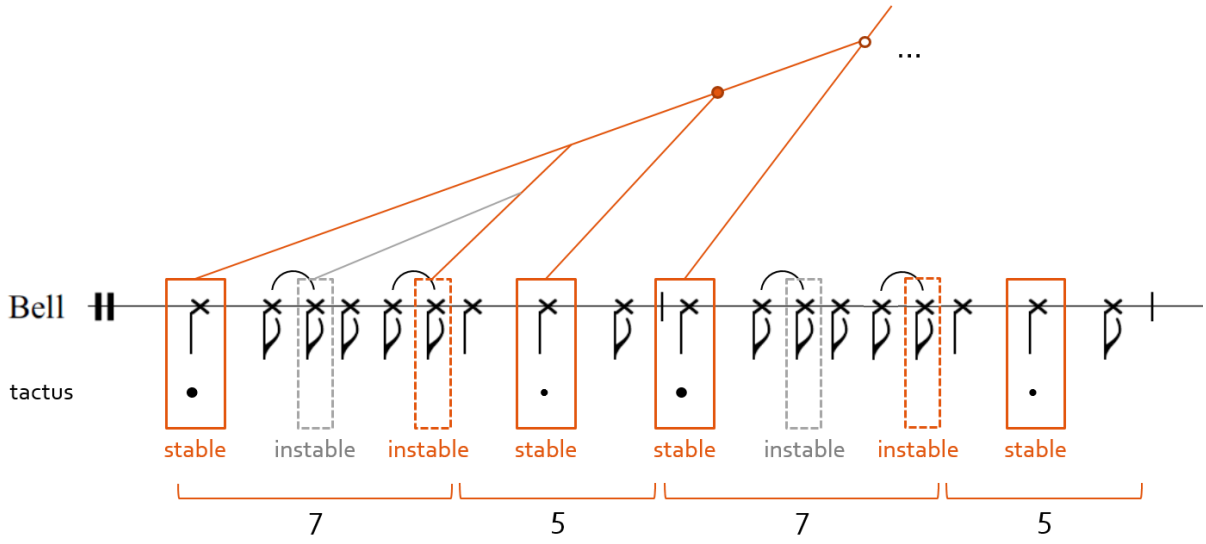
Further, the hemiola created by clapping adds more fine-grained tension-relaxation patterns in the middle of the metrical cycle. The second and third attack of the clap co-occur with those of the bell on weak beats in the metrical framework of Agbekor (gray beats marked with red arrows in Figure 12.8a). This creates a sense of two-beat periodicity in contrast to three-beat periodicity which is represented in the primary beats. That is, in the first half, grouping and metrical structure do not align and grouping structure is even in support of a conflicting metrical interpretation with a two-beat periodicity. Such a temporal conflict causes a phenomenon called "metrical dissonance" (Krebs, 1999; London, 2012a) and causes maximal

instability in terms of meter. However, in the second half, the attacks of the clap completely align with the primary beats and thus grouping structure conforms to metrical structure. As the attack of the clap corresponds to the third beat in the tactus[44] (marked with a red dashed square in Figure 12.8a), it is more stable than the second beat in the tactus, while it is less stable than the first and the last beat. The updated tension-relaxation pattern of the standard pattern is represented in Figure 12.8b. In contrast to Figure 12.7, the branching of this hierarchical structure additionally considers 7+5 grouping structure and different degrees of instability.

a)



b)



[44] In the current example, the primary beats correspond to the foot movements of the dancers.

Figure 12.8 Tension-relaxation pattern of the standard pattern of Agbekor (final). a) Stability of the primary beats by taking hemiola displayed in the claps into account. b) Updated prolongational tree representing tension-relaxation pattern of the standard pattern of Agbekor.

Formal mathematical models integrating meter and grouping to encode affect are still missing. Toussaint (2003) focused on combinatorial aspect of the seven-stroke 12/8 rhythm in African and Afro-American music and claimed that there are mathematical constraints for rhythmic realization. Although 12!/(7!)(5!), i.e., 792, patterns are theoretically possible, actual rhythmic patterns are limited to a subset. Most of those theoretically possible rhythmic patterns are not 'good enough' to serve as a timeline and only ten out of 792 possible patterns are known to be used in traditional music (Toussaint, 2003). Those ten patterns notated by using the interval-vector notation are listed in the *Table 12.1* and named after the convention used in Toussaint (2003). First of all, he showed that all ten patterns can be derived from three canonical patterns (see also Figure 12.9 I - III). Second, the pattern called "Ashanti" is a Euclidean rhythm[45] which can be built by the Bjorklund algorithm and has the property that the onsets are maximally evenly distributed (Toussaint, 2005). The rest of patterns belonging to the canonical pattern III can be then created by rotating Ashanti pattern. Finally, the remaining four patterns can be obtained by permutations of the interval-vector elements of Ashanti pattern, i.e., two-times one and five-times two. Thus, all ten rhythms belong to the same interval combinatorial class. However, it is still not clear why only ten of (7!)/(2!)(5!), i.e., 21, possible permutations occur in world music culture.

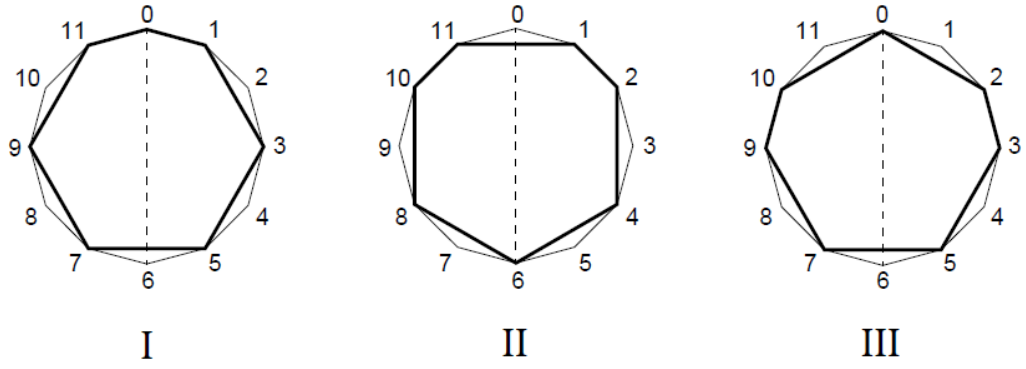


Figure 12.9 Three canonical patterns generating ten African timeline patterns. Adopted from Toussaint (2005, p. 5) with permission by Godfried Toussaint.

[45] Euclidean rhythms are rhythms which have "the property that their onset patterns are distributed as evenly as possible" (Toussaint, 2005, p. 1).

Table 12.1 Ten African timeline patterns notated by using the interval-vector notation.

Canonical	Name	Pattern	Max. even?	Off-beatness
I	Sorsonet	1+1+2+2+2+2+2	No	1
II	Asaadua	2+2+2+1+2+1+2	No	1
II	Soli	2+2+2+2+1+2+1	No	1
II	Tonada	2+1+2+1+2+2+2	No	1
III	Ashanti	2+1+2+2+1+2+2	Yes	2
III	Bemba	2+1+2+2+2+1+2	Yes	2
III	Bembé-2	1+2+2+1+2+2+2	Yes	2
III	Tambú	2+2+2+1+2+2+1	Yes	2
III	Yoruba	2+2+1+2+2+1+2	Yes	2
III	Bembé	2+2+1+2+2+2+1	Yes	3

Although Toussaint's approach makes it possible to describe characteristics of rhythmic patterns in a wide range of musical cultures, it does not tackle the computational problem of music, namely linking sound and affect. As discussed above, tension-relaxation patterns of non-tonal music emerge from the relationship between hierarchical metrical structure giving rise to metrical accents and phenomenal accents created by the onsets at the surface. Further, grouping structure adds refinements to the so-created tension-relaxation patterns. However, Toussaint's theory considers neither hierarchical metrical and grouping structure nor affect. Toussaint (2005) characterizes African rhythms as off-beat[46] rhythms, i.e., patterns with the onsets at the off-beat positions including 1st, 5th, 7th, and 11th beats in 12 beat cycle (see Figure 12.10). Those positions are off-beat because they do not belong to the beats included in binary and ternary subdivision of 12 beats. For example, Bembé is maximally off-beat because three of those off-beat positions are included in its onsets. This measure considers binary and ternary metrical organization, but does not touch the dynamics of the tension-relaxation patterns.

[46] Here, the off-beats are weak beats concerning both binary and ternary subdivisions. That is, for example, the beat 3 and 9 are not classified as off-beats even though they are off-beats in the binary subdivision. This is because they are strong beats in the ternary subdivision.

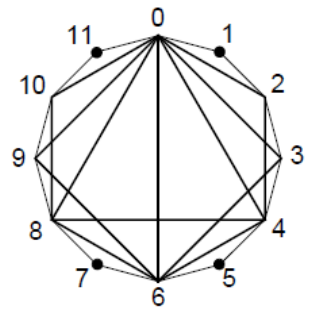


Figure 12.10 African rhythms as offbeat rhythms. Adopted from Toussaint (2005, p. 11) with permission by Godfried Toussaint.

Nevertheless, research on rhythmic syntax can learn an important lesson from the mathematical approach of Toussaint (2005), namely a cyclic organization of beats. As shown in Figure 12.9, the rhythmic patterns of African drum ensemble music discussed above are organized in 12 beat cycle[47]. Further, meter in a wide range of musical cultures such as Western, African, and Hindustani music, builds on such a cyclic organization with diverse cardinality, i.e., the total number of beats involved in a cycle (London, 2012a). Patel (2008, p. 96) suggests periodicity, i.e., "a pattern repeating regularly in time", as a central aspect of musical rhythm, which differs from speech rhythm. This characteristics of musical rhythm can be best yielded by cyclic organization of beats. Thus, the cyclic organization seems to be one candidate principle in rhythmic syntax explaining why musical rhythm is the way it is, e.g., featuring periodicity.

12.3 Computational principles of rhythmic syntax

As discussed in PART II, recursion and hierarchical structure building are two core concepts of syntax at the computational level. Concerning musical rhythm, metrical and grouping structures are organized recursively and hierarchically. Metrical structure is built up by embedding beats into a beat and grouping structure is constructed by embedding groups into a group. Those hierarchical structures are yielded by binary and ternary combinatorics. A set of rules (e.g., MWFRs, GWFRs, MPRs, and GPRs) constrain[48] combinatorial possibility. Moreover, once we consider a hierarchical structure encoding tension-relaxation patterns, binary (but not ternary) combinatorics becomes more important. The central aspect of such a

[47] In this case, "beat cycle" refers to the beats at the lowest subdivision level.
[48] I use here the term "constrain" instead of "determine" as the rule system introduced by Lerdahl and Jackendoff (1983) are not generative rules deriving grammatical sequences (see also Rohrmeier & Pearce, 2018). Their rules rather constrain combinatorial possibilities.

hierarchical structure is headedness, i.e., the most stable event is considered as a head which is determined by means of the relationship between metrical and phenomenal accent and refined on the basis of grouping structure as discussed above. In this way, the current chapter investigated the computational problem for rhythmic syntax as linking sound and affect by mapping metrical and grouping structure and temporal sequence.

In constructing a headed hierarchy, binary combinatorics identifies stability relationship between events unambiguously, but ternary combinatorics does not. Lerdahl and Jackendoff (1983) also pointed out the potential for strictly binary combinatorics in music. It is possible that 3 is additionally segmented in 2+1 (e.g., half note + forth note) or 1+2 (e.g., quarter note + half note) (Lerdahl & Jackendoff, 1983, pp. 327–330). Concerning Mozart's Sonata K. 331 in 6/8 measure, three eighth notes are combined in 2+1, i.e., one quarter note (= two eighth notes) and one eighth note, as particularly seen in the accompaniment. This is in line with a study conducted by Palmer and Krumhansl (1990) showing that events occur more often at the first, third, fourth, and sixth eighth note in 6/8 measure, indicating (2+1)+(2+1) structuring. In the timeline pattern of West African drum ensemble music discussed above, three eighth notes are combined into 1+2, i.e., one eighth note and one quarter note (= two eighth notes), as well as 2+1, i.e., one quarter note (= two eighth notes) and one eighth note. In addition, there is a tendency to prefer binary metrical interpretation of sequences (Huron, 2006; London, 2012a). Thus, it is likely that the first two incoming events are put together [x, x] foremost and then the third event is integrated [[x, x], x].

A further aspect playing a central role in rhythmic syntax is a cyclic organization of beats giving rise to periodicity in musical rhythm. Cyclicity as a structuring principle also earned considerable attention in language research (Boeckx, 2014; Boeckx & Theofanopoulou, 2018; Murphy, 2015; Samuels, 2011). In general, cyclicity refers to recurrence of events at regular intervals. In its simplest form, cyclicity is realized as isochronous beats in musical rhythm. However, musical rhythm is not a mere line up of regular recurrent events, but a structured sequence yielded by combining events hierarchically, giving rise to (isochronous or non-isochronous) metrical cycles. Importantly, once events are combined in a unit, only one event, namely the head of each unit, is visible to the next combinatorial cycle. Once those units are combined in a larger unit, again, only one event is visible to the next cycle, and so on.

Hierarchical combination of a head and an elaboration event in a cyclic manner (e.g., {{head, elaboration}, {head, elaboration}}[49]) gives rise to musical rhythm.

One relevant syntactic approach considering both combinatorial and cyclic computations was proposed by Knott (2014) in research on sensorimotor syntax. The central idea which is relevant to rhythmic syntax is one regarding attentional and motor actions as 'building blocks' to plan sensory-motor sequences. According to Knott (2014), sensorimotor sequences are hierarchically planned by combining attentional and motor actions recursively. Attentional and motor actions are *deictic operations*, i.e., "cognitive operations which bring about updates in the agent's physical relationship with the environment, and also in his internal cognitive representations" (Knott, 2014, p. 9). Thus, as seen in the following example of a cup-grapping episode (*Figure 12.11*), a deictic operation (e.g., attend_man) performed in a particular context (e.g., c1) causes a new context (e.g., c2) according to which the subsequent deictic operation (e.g., attend_cup) is performed. This combinatorial procedure cyclically continues until a motor program (e.g., *grap* motor program) is activated. In addition, Knott (2014) suggest a hypothesis that a hierarchical structure as illustrated in *Figure 12.11* is a working memory representation of a cup-grapping episode.

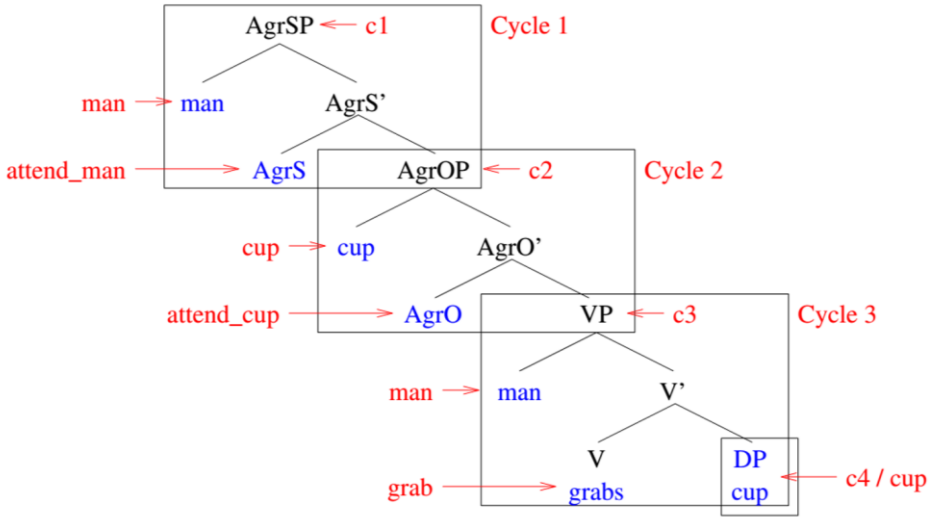


Figure 12.11 Cyclic organization of sensorimotor syntax. The figure is adopted from Knott (2014, p. 24), *Biolinguistics*, 8: 001-052.

The hierarchical structure represented in Figure 12.11 parallels the X-bar schema developed in linguistic research. The hierarchical structure consists of four phrases (XP), i.e.,

[49] Here, I use "{ }" notation as both head-elaboration and elaboration-head sequential order are equally possible. See also footnote 43.

a subject agreement phrase (AgrSP), an object agreement phrase (AgrOP), a verb phrase (VP), and a determiner phrase (DP). In linguistic X-bar tree structure, the VP represents a predicate-argument structure[50] and subject and object agreement features are checked within the agreement phrases. Knott (2014) proposed to apply this linguistic structure to sensorimotor syntax. In sensorimotor syntax, VP is interpreted as representing a motor program and AgrPs as representing attentional action. Each phrase corresponds to a cycle of context-updating. As such, this model relies too much on linguistic sentential structure and is difficult to apply to rhythmic syntax. However, at a very abstract level, the concepts introduced in Knott (2014) might be applicable to rhythmic syntax.

On the basis of Knott's (2014) conception of a sensorimotor syntax, rhythmic syntax can be understood as combinatorial and cyclic computation of deictic operations such as attentional and motor actions, in which meter can be understood as attentional action and drumming, hand clapping, and dancing are motor actions. In this line, for example, London (2012a, p. 91) suggested that "[a] meter is a coordinated set of periodic temporal cycles of sensorimotor attention". In particular, London (2012b) suggested primary beats as peaks of attentional energy by referring to Edward Large's and Mary Ries Jones' dynamic attending theory (e.g., Large & Jones, 1999). In addition, cyclic computation of rhythmic syntax has an important implication for memory: Cyclic computation achieves minimization of memory load by keeping a minimum amount of information accessible for each manipulation step in hierarchical structure building. This is similar to chunking, an information processing mechanism proposed by Miller (1956), in terms of optimizing memory capacity. In general, a "chunk" is "a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet et al., 2001, p. 236). Thus, hierarchical structure of musical rhythm generated by combinatorial and cyclic computation may be regarded as working memory representation in parallel to the suggestion by Knott (2014). In this line, Palmer and Krumhansl (1990) showed that people memorize a probe tone better if it is on the strong beat, i.e., the head.

Last but not least, hierarchical combination of a head and an elaboration event in a cyclic manner is a general principle for temporally structured cognitive capacities including speech, music, and action (Asano & Boeckx, 2015; Fitch & Martins, 2014; Lashley, 1951; Lerdahl & Jackendoff, 1983). As Lerdahl and Jackendoff (1983, p. 330) suggested, this is "more highly structured than one articulated only in terms of "chunking", since it accords one

[50] In the current case, predicate, i.e., the verb "grab", takes two arguments, i.e., object "cup" and subject "man".

element of each "chunk" the privileged status of head". In other words, heads are needed to manipulate chunks hierarchically. In addition, as I will discuss in details in Section 15.1 (p. 146), prosodic hierarchy consists of onset, nucleus, and coda with the nucleus as a head and action hierarchy is consisted of preparation, head, and coda. They parallel hierarchical structures of musical rhythm and dance. Therefore, hierarchical and cyclic combination of events are general principles of temporally organized cognitive domains.

13 Rhythmic syntactic processing and its neural implementation

13.1 Processing structural relationships in musical rhythm

Beat-based encoding of rhythm can be regarded as hierarchical processing in parallel to that of tonal-harmonic processing as discussed in PART II. In processing musical rhythm, auditory events are perceived in relation to the primary beats which act as mental reference points. Thus, beat-based encoding of rhythm, first, entails a sub-process abstracting the primary beats from the musical surface (i.e., beat abstraction). Moreover, to establish structural relationship between auditory events based on the primary beats, listeners internally generate primary beats to produce expectancies in real-time (i.e., beat generation). This also enables the listeners to recognize structural violations. In the following section, I claim that both beat abstraction and beat generation can be considered as hierarchical processing, i.e., mapping between hierarchical structure and temporal sequence. I regard "beat-based encoding of rhythm" (PART I) as an umbrella term[51] referring to the interaction of both sub-processes to process structural relationship in musical rhythm. In addition, I consider those aspects of beat-based encoding of rhythm which relate to hierarchical combination of elements into sequence as belonging to rhythmic syntactic processing.

Although there is skepticism about the existence of metrical hierarchy extending beyond the primary beats (e.g., Vuust & Witek, 2014), there are some arguments for the need of metrical hierarchy in beat-based encoding of rhythm. First, several tapping studies showed that, at fast tempi, people tend to choose the tactus at a higher level in the metrical hierarchy than at slow tempi (McAuley, 2010 and references therein). This systematic relationship between the tactus level in metrical structure and tempo is in favor of hierarchical processing. It also provides an additional argument for assuming hierarchical metrical structure in beat-

[51] In music cognition research, other terms such as "beat induction" (Honing, 2012) and "beat perception" (Patel, 2006) are used instead. To avoid invoking connotations which are sometimes associated with those terms, I use "beat-based encoding of rhythm" in the current thesis. Honing (2012 in footnote a) and Fitch (2013, p. 9) discuss some terminological issues.

based encoding of rhythm. The discussion on the relationship between hierarchical metrical structure and tempo opens up possibility to link hierarchical processing in musical rhythm to more general processing principles underlying temporal organization.

McAuley (2010) suggested that the influence of tempo on listener's choice of the tactus level in metrical hierarchy relates to the preferred tempo which lies around 600 ms inter-beat interval (IBI) or inter-tap interval (ITI), i.e., 100 beats per minute (BPM), regardless of perceptual or motor tasks.[52] At this tempo, the motor-related circuitry including the supplementary motor area (SMA), basal ganglia, and premotor cortices, together with additional areas, is more activated than at a slower tempo (McAuley, Henry, & Tkach, 2012). The preferred tempo is optimal for binding sensory and motor information into a sensory-motor unit as well as integrating sensory-motor units (e.g., the primary beats) into a memory unit (e.g., measure). For example, the preferred tempo corresponds to the processing level mediating external sensory data and internal cognitive programs, namely the embodiment level (Ballard, Hayhoe, Pook, & Rao, 1997).[53] More precisely, as the embodiment level by Ballard and colleagues (1997) refers to saccadic eye movements taking place in the 300 ms time window, the limb movements in the 600 ms span are between the embodiment and cognitive level.[54] At the embodiment level, sensorimotor primitives are created in a fraction of a second to bind the current items into memory. They are then synthesized to execute the task.

In addition, given a quarter note has the duration of 600 ms in 4/4 measure and is perceived as the primary beats, a half note corresponding to the metrical level which is one level above equals to 1,2 s. The entire measure, then, is 2,4 s of duration, which approximately parallels the '3 second window' of automatic, pre-semantic temporal integration proposed by Pöppel (2004, p. 298) and matches the time window between auditory echoic memory (approx. 1 second) and auditory short-term memory (approx. 4-8 seconds) (Snyder, 2016, p. 168).

In any case, it deals with temporal hierarchy, i.e., nesting of time-spans into a time-span. A direct mapping hypothesis between such a temporal hierarchy and nested neural oscillations was suggested by different authors working in the framework of dynamic attending theory (for reviews, see Mari Riess Jones, 2016; Large, 2008). Neural oscillations, generated by balanced interaction between excitatory and inhibitory neural activities, are considered to be

[52] Preferred tempo for children between the ages of 4 and 7 years lies between 300 and 400 ms (McAuley, 2010).
[53] It is also worth noting that Knott (2014) relates his sensorimotor syntax to the embodiment level of Ballard and colleagues (1997).
[54] Ballard and colleagues (1997) introduced two neural levels (1 ms for neuron spike and 10 ms for lateral inhibition in neural circuit) as well as attentive level (50 ms for deliberate act) in addition to embodiment and cognitive level.

hierarchically organized from fast to slow oscillations (e.g., delta 0.5–1.5 Hz < theta 4–10 Hz < alpha 8–12 Hz < beta 13–30 Hz < gamma 30–120 Hz) (Buzsáki, Logothetis, & Singer, 2013). Faster oscillations are considered to be embedded into slower oscillations. The dynamic attending theory assumes that multiple, self-sustaining neural oscillations are activated by different time levels within metrical structures, indirectly relating the hierarchy in neural oscillations and hierarchical metrical structure (Mari Riess Jones, 2016). Jones (2016) extended the dynamic attending theory and introduced the metric binding hypothesis stating that simultaneously active oscillations at different hierarchical levels internally entrain overtime to internalize metrical accents.

Second, hierarchical processing can be investigated by examining syncopation, a phenomenon well-known in musicology. Figure 13.1a and c are syncopated, whereas Figure 13.1b is unsyncopated. The difference between those sequences is, roughly speaking, that in the former case "a "heavier" note is tied back to a "lighter" sounded note" and in the latter case "the "heavier" note is the first of the tied pair" (Longuet-Higgins & Lee, 1984, p. 430). The weight of a given note or rest can be determined by their position in the underlying metrical hierarchy (Figure 13.1). Because of the conflict between the internal representation (hierarchical metrical structure) and the information from the actual rhythmic sequence, syncopation is considered to be complex and requires more cognitive resources (Fitch & Rosenfeld, 2007). Based on the note weight, syncopation strength can be calculated as the weight of R (a rest or a tied note) - the weight of N (the next sounded note before R) (Longuet-Higgins & Lee, 1984).[55] The syncopation index of the whole sequence is the sum of the individual syncopation strengths (Fitch & Rosenfeld, 2007).

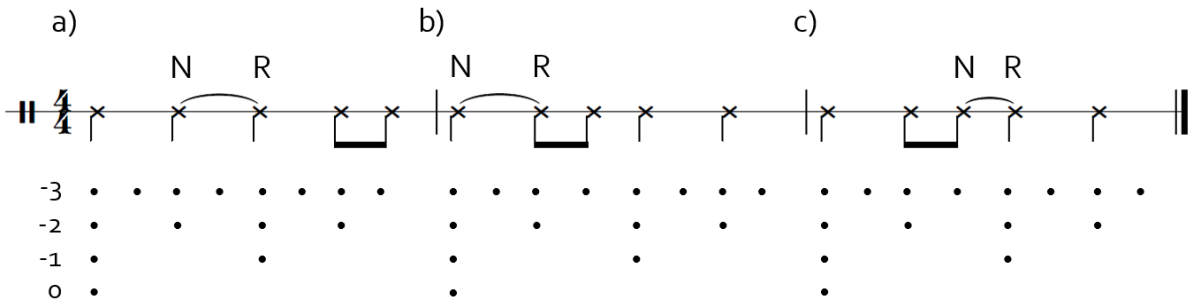


Figure 13.1 Syncopated rhythms with different degree of syncopation index.

[55] In a complementary model which assumes a generation of an "internal clock" (i.e. meter and the beat) specifying the temporal structure in the pattern, too, the complexity of rhythmic sequences is mainly determined by the strength of the beat induction (also called "clock induction"), measured on the basis of the relationship between the best fitting clock and the perceived accent of the sequences (Povel & Essens, 1985).

For example, the strength of the syncopated rhythm (Figure 13.1c) is calculated as following: -1 (the weight of R) - -3 (the weight of N) = 2. Because there is only one syncopation in this rhythmic sequence, the syncopation index of the whole sequence is 2. The higher the syncopation index, the more complex the rhythmic sequence (Fitch & Rosenfeld, 2007). The other syncopated rhythm (Figure 13.1a) has syncopation index of 1 (i.e., -1 - -2), meaning that this rhythm is less complex than the former rhythm (Figure 13.1c). The remaining rhythm (Figure 13.1b) has a syncopation index of -2 (i.e., -2 - 0), meaning that the rhythm does not form a syncopation. In this way, complexity of the rhythmic sequence can be identified on the basis of hierarchical structure.

In line with this theoretical consideration, Fitch and Rosenfeld (2007) showed that tapping along the primary beats with the syncopated rhythms as well as reproducing and recognizing syncopated rhythms are significantly harder in comparison with the same task with unsyncopated rhythms. The higher the syncopation index, the worse the task performance. Moreover, in the same experiment, they found that participants tend to 'reset' the internally represented beat during tapping along syncopated rhythms, i.e. they preferred unsyncopated representations of rhythmic sequences. This is also in line with the theoretical consideration that 'natural' interpretations of rhythmic sequences are those that enable listeners to interpret them as a realization of an unsyncopated passage (Longuet-Higgins & Lee, 1984).

In addition, a recent study reported a significantly larger Mismatch Negativity (MMN) amplitude when a deviant sound (i.e. the omission of a sound) occurred in strong metrical positions (indicating that these sounds were more unexpected), compared to when it occurred in weak metrical positions (Bouwer, Van Zuijen, & Honing, 2014). The omission in the strong metrical position created a syncopation. The MMN is taken to reflect the processing of expectancy violations, and its amplitude scales with the degree of the expectancy violation (Honing, Bouwer, & Haden, 2014; Winkler, 2007). Thus, the result that a beat omission in strong metrical position causes a greater expectancy violation and elicits a larger MMN is in favor of hierarchical processing in music in terms of mapping hierarchical metrical structure and rhythmic sequence. Moreover, another EEG experiment showed that a syncopated ending in contrast to an unsyncopated ending elicits an early right anterior negativity (ERAN), i.e., an ERP component associated with music syntactic violation (Sun, Liu, Zhou, & Jiang, 2018).

Importantly, syncopation is possible just because there is hierarchical metrical structure that is cyclic and highly regular. Without relating the internally represented hierarchical metrical structure and the rhythmic sequence, one cannot tell whether some rhythm is syncopated or not. Moreover, the syncopation strengths and their behavioral correlates

discussed above can be explained only if we acknowledge hierarchically structured representations underlying the rhythmic sequences. Indeed, the syncopation strengths are constructs yielded by the mapping between hierarchically structured metrical representations and rhythmic sequences.

Finally, even in hearing an acoustically identical series of tones, listeners tend to subjectively hear patterns of two or three, i.e. they hear some auditory beats as more important than the others (Fitch, 2013; London, 2012a). Two ERP studies showed that a late positive P3-like component was larger if deviants (i.e., 4 dB softer in comparison with the standard tones) were placed at (putatively) subjectively metrically strong positions (i.e., odd numbered tones) of isochronous unaccented stimulus than at subjectively metrically weak positions (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Potter, Fenwick, Abecasis, & Brochard, 2009). Those experiments also provided evidence that people tend to form subjective accentuation in binary meter. A MEG study tested participants imagining march or waltz metrical structure while listening to an isochronous unaccented stimulus and showed that beat-related beta-power[56] decrease associated with the imagined strong beat (i.e., down-beat) was larger than that associated with the imagined weak beat following the imagined strong beat (Fujioka, Ross, & Trainor, 2015).

13.2 Processing structural ambiguity

Any given note sequence can be interpreted in potentially infinite ways and is therefore in principle infinitely ambiguous (Longuet-Higgins & Lee, 1984), but an ‘experienced listener’ ‘knows’ that there are a few particular ways to infer structures from note sequences adequately (Lerdahl & Jackendoff, 1983). In this case, ambiguity means that there are multiple possible well-formed representations. For example, Mozart’s Sonata KV 331 can be interpreted in form of different metrical structures represented as Figure 13.2a and b, but not in the way represented in Figure 13.2c. That is, the structural representations in Figure 13.2a and b are well-formed, but that of Figure 13.2c is not. However, experienced Western listeners would intuitively choose the interpretation of Figure 13.2a even though according to Lerdahl and Jackendoff (1983) both structural representations are possible interpretations. That is, the representation in Figure 13.2a is more preferred by Western listener than that in Figure 13.2b. The choice of

[56] EEG signals recorded on the scalp can be decoded into different frequency bands of neural oscillations, i.e., synchronized rhythmic patterns of electrical activity induced by synchronized activations of large amount of neurons (in this case). The beta-band is such a frequency band corresponding to 10-30 Hz oscillations. The power of each frequency band in EEG signals are examined by time-frequency analysis.

well-formed and preferred interpretations is not arbitrary, but there are rule systems and constraints which determine well-formedness and preference of structures (Lerdahl & Jackendoff, 1983; London, 2012a).

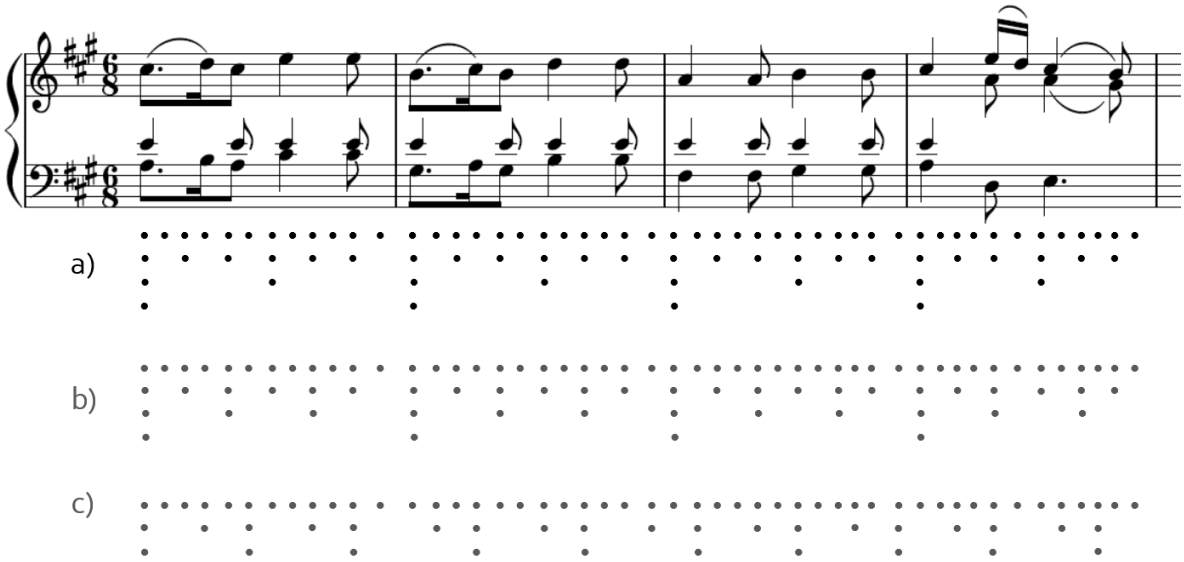


Figure 13.2 Well-formed and ill-formed metrical structures of Mozart's Sonata KV 331.

For example, metrical well-formedness rules (MWFR) (see also (12-1) in the current thesis) state how to yield a metrically well-formed representation in parsing a musical sequence. I discussed MWFRs and the others concerning computational-representational theories of rhythmic syntax, but, at the same time, I agree with Rohrmeier and Pearce (2018) that GTTM is a parsing model. Thus, the rule systems of GTTM rather constrain amounts of representations available to the parser and weight preferred representation in online processing. In case there are conflicting well-formed representations, like in Figure 13.2a and b, metrical preference rules (MPR) take effect. In the case of the above-mentioned example, the representation Figure 13.2a 'wins' on the basis of the MPR 3: "[p]refer a metrical structure in which beats of level L_i that coincide with the inception of pitch-events are strong beats of L_i" (Lerdahl & Jackendoff, 1983, p. 347).

Similarly, metrical structure of the Agbekor bell pattern can be represented as a) or b) (Figure 13.3). Western listeners would prefer Figure 13.3a and extract the primary beats at each quarter or half note, but for African listeners the normal interpretation is Figure 13.3b with the primary beats at every dotted quarter note. Again, both interpretations are well-formed, but one of them is more preferred over the other. For example, MPR 10 "[p]refer metrical structures in which at each level every other beat is strong" (Lerdahl & Jackendoff, 1983, p. 348) is in favor

of the metrical structure Figure 13.3a. This is also in line with the binary meter bias in Western music (Huron, 2006). However, polyrhythmic textures created by Agbekor drum ensemble including dance and song prefer the metrical structure of Figure 13.3b and even a non-isochronous meter such as Figure 13.3c. Although the representation Figure 13.3c violates the MWFR 4, it can be considered as well-formed on the basis of the updates called well-formedness constraints (WFC) introduced by London (London, 2012a) to extend the research to non-Western culture: "WFC 4.2.2 If the beat cycle is non-isochronous, then either (1) it is maximally even, or (2) the cycle above the beat cycle, in most cases the half-measure cycle, must be maximally even" (London, 2012a, p. 158).

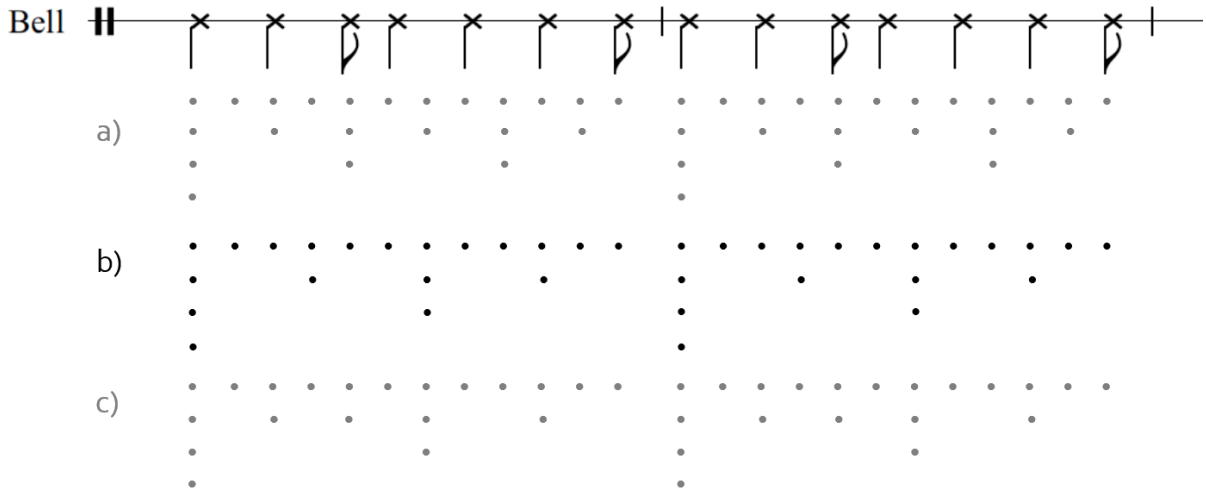


Figure 13.3 Well-formed metrical structures of Agbekor's standard pattern.

Importantly, each metrical structure creates different tension-relaxation patterns. Regarding Figure 13.3a with the tactus at the half-note level, the emerging stability pattern is stable → stable → unstable and thus the tension-relaxation pattern has the form of weak prolongation → tension. As discussed already, the metrical structure Figure 13.3b creates stable → unstable → unstable → stable pattern, meaning tension → relaxation. Concerning Figure 13.3c with the tactus at the second level (i.e., 2-2-2-3-3 or short-short-short-long-long), the stability pattern is in form of stable → stable → stable → unstable → stable which creates the tension-relaxation pattern of weak prolongation → strong prolongation → tension → relaxation. Moreover, throughout the performance there is usually one prominent metrical representation although both Figure 13.3b and Figure 13.3c are preferred metrical representations of Agbekor and the prominent metrical representation changes from section to section depending on the currently dominant rhythm. For example, Agbekor dance foregrounds the metrical structure

Figure 13.3b and the song mainly the metrical structure Figure 13.3c. In addition, by putting phenomenal accent at different positions with different instrument, additional tension-relaxation patterns can be created. In this way, Agbekor performance seems to make use of ambiguity to create diverse tension-relaxation patterns and a large-scale structure.

Another kind of ambiguity, i.e., the existence of some conflicting representations, is reflected in "rhythmic garden-path" effects in which the established metrical representation turns out to be implausible at a later time point, i.e., the information provided at a later time point conflicts with the established metrical representation (Slevc & Okada, 2015). Based on the example in the Figure 13.4, Slevc and Okada (2015) suggested that listeners first establish either binary meter (Figure 13.4a) or ternary meter (Figure 13.4b) and, in the former case, potentially revise the first interpretation after having reached the 4th bar. Similarly, in Figure 13.5, metrical ambiguity inherent in the first few notes is resolved by later notes (Longuet-Higgins & Lee, 1984). In this example, in the first two bars, it is not clear whether underlying metrical structure is in form of Figure 13.5a or b. Once the parser reached the third bar, a clear preference for each structure can be established on the basis of MPR 3. In this case, too, depending on which metrical representation was built in the first two bars, the listener encounters the situation in which the established metrical representation and the alternative more plausible one are in conflict.

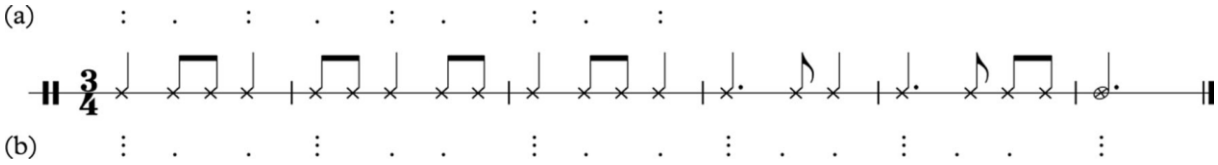


Figure 13.4 Rhythmic garden-path. The figure is adopted from Slevc and Okada (2015, p. 643), Psychon Bull Rev, 22 (3): 637–652, with permission by Springer.

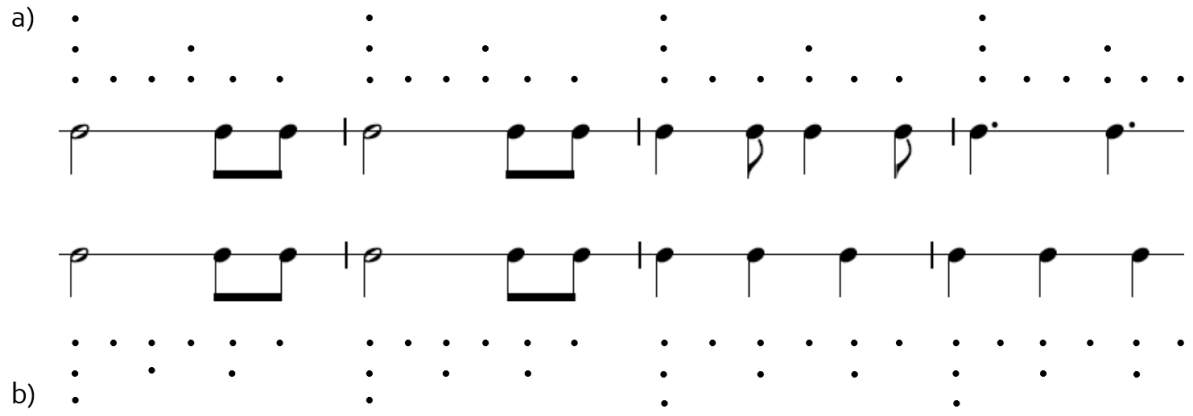


Figure 13.5 Ambiguity resolution in rhythm processing as discussed by Higgins and Lee (1984).

Slevc and Okada (2015) pointed out that it is not clear whether the listeners resolve the conflict. In doing so, they refer to a study conducted by Vazan and Schober (2004) which demonstrated that only a few participants revised their initial metrical interpretation. Vazan and Schober (2004) used the rock song "Murder by Numbers" by The Police which contains metrical ambiguity. This song begins with a clear ternary meter represented in the percussion and then 'changes' to a binary meter. For the remaining of the piece, the binary meter is more dominant. In the experiment, the participants first tapped the meter which they found most natural during listening the entire piece. After this familiarization phase, the first 80 seconds of the piece with the switching from ternary to binary meter was presented repeatedly. One group was instructed to tap the meter which they found most natural, while the other group was instructed to tap in a single coherent way. During the first listening, most participants switched their tapping from ternary to binary meter. Even after repeated listening, most participants tended to tap the ternary meter first and then change to the binary meter. Only a few participants in the coherent tapping group showed increasing tendency to tap only binary meter from the beginning. This result was interpreted as the listeners not reinterpreting metrical structure, i.e., not using retrospective listening strategy.

However, this particular experiment cannot be used as an evidence against garden-path processing in musical rhythm. First, the fact that the participants spontaneously switch from ternary to binary meter parallels the garden-path effect. The switching from ternary to binary meter, i.e., from SW<u>W</u> to SW<u>S</u>W,[57] engages a reinterpretation of the underlined weak beat as the underlined strong beat. That is, structurally, the listeners should switch from [[x, x], x]] to [[x, x], [x, x]]. Second, even in the garden-path sentences, the reinterpretation often takes place in a local manner. Let us consider the following example garden-path sentence (13-1).

(13-1) The man accepted the money could not be spent yet.

In this example sentence (13-1), the noun phrase "the money" is first interpreted as the direct object of the verb "accepted", but then is reinterpreted as the subject of the passive voice subordinate clause. Thus, the syntactic function of the noun phrase "the money" is reinterpreted and the interpretation of something accepted by the man, but neither the syntactic structure of "The man accepted [something]" nor its meaning is affected by the garden-path effect. This indeed parallels the finding by Vazan and Schober (2004) that the later metrical interpretation

[57] S = strong beat, W = weak beat

does not affect the early metrical interpretation. The rock song “Murder by Numbers” makes use of the metrical ambiguity to modify the tension-relaxation pattern of each section by foregrounding different metrical structure trough phenomenal accents created by the percussion. Thus, it is appropriate that the listeners tend to experience the switching between ternary and binary meter. Similar switching between ternary and binary meter is engaged in processing non-isochronous (additive) meter or hemiola. A MEG study that investigated processing of hemiola (3+3+2+2+2) demonstrated that auditory evoked response change from ternary to binary pattern just before the point where metrical change takes place (Fujioka, Fidali, & Ross, 2014). This study thus provides an additional evidence for switching metrical representations.

13.3 Processing affect encoded by rhythmic syntax

As repeatedly pointed out, hierarchical structure of music represents stability relationship that encodes affect, i.e., pattern of tension and relaxation. In a framework considering beat-based encoding of rhythm in terms of hierarchical processing, tension is created by conflict between hierarchical metrical structure and phenomenal accent or group boundary. The examples are syncopation and polyrhythm. The processing of affect in musical rhythm can be investigated in terms of a parallel parser that keeps alternative structural analyses, and operates on the winner-take-all principle. A parser concatenates incoming events, inserts multiple possible structural analyses to the part of music heard so far, and weights one analysis as the most likely representation (Jackendoff, 1991). In the course of processing, however, a parser might encounter a situation in which the established representation is in conflict with the incoming information from the musical surface or multiple representations are equally likely. The former situation fits to syncopation, while the latter to polyrhythm. Such situations are then processed as instable and causing tension. A parser ‘projects’ the established representation beyond the heard part of music. Thus, tension can be investigated in terms of expectancy violation or prediction error as suggested within different frameworks (Huron, 2006; Jackendoff, 1991; L. B. Meyer, 1956; Narmour, 1977; Vuust et al., 2014; Vuust & Witek, 2014).

In addition, musical affect has very strong relation to movement and bodily representation (Bierwisch, 1979; Colling & Thompson, 2013; Jackendoff, 1987; Molnar-Szakacs & Overy, 2006; Trost & Vuilleumier, 2013; Witek, Kringelbach, & Vuust, 2015). Based on the well-known experimental cartoon by Heider and Simmel (1944) showing that the character of motion can ascribe animacy to the geometric figures and encode affect, Jackendoff

and Lerdahl (2006) argue that temporal patterns can similarly convey affect in music. I therefore suggest that temporal patterns in music convey affect via engagement of motor system. Just listening to music activates several motor areas in the brain (for reviews, see Cameron, Pickett, Earhart, & Grahn, 2016; Grahn, 2012; Leow & Grahn, 2014; Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015). Body movement influences one's way to abstract metrical representation (Phillips-Silver & Trainor, 2005, 2007). The strong link between music and dance also supports the connection between music and motor system (Fitch, 2016; Jackendoff & Lerdahl, 2006). Moreover, the music-dance link is a key domain to investigate how affect is processed in music via engagement of motor system. Especially, entrainment and groove are two relevant phenomena.

In general, "entrainment" refers to "the process by which independent rhythmical systems interact with each other" (Clayton, 2012, p. 49) and can be observed not only in musical context, but also in several biological, physical, and social contexts. Trost and colleagues (2017) suggested that different levels of entrainment such as neural, perceptual, motor, and social entrainment all could induce affect in the subjects sharing a musical activity and thus affective entrainment takes place at all levels and also through interdependency of those levels. Groove is characterized as "a pleasurable drive toward body-movement in response to rhythmically entraining to music" (Vuust & Witek, 2014, p. 9) and is thus "[o]ne common source of affective entrainment in music" (Witek et al., 2015). Groove is often discussed in terms of prediction error (Vuust et al., 2014; Vuust & Witek, 2014), but the factors that lead to groove are not limited to this (Fitch, 2016; Merker, 2014). Further, entrainment and groove are often discussed in embodiment frameworks and are not investigated within a syntactic framework. However, entrainment presupposes the existence of regular beats that, as discussed above, are abstracted from musical surface. Moreover, social entrainment presumes the presence of shared beats between individuals, indicating that there should be common principles for beat abstraction.

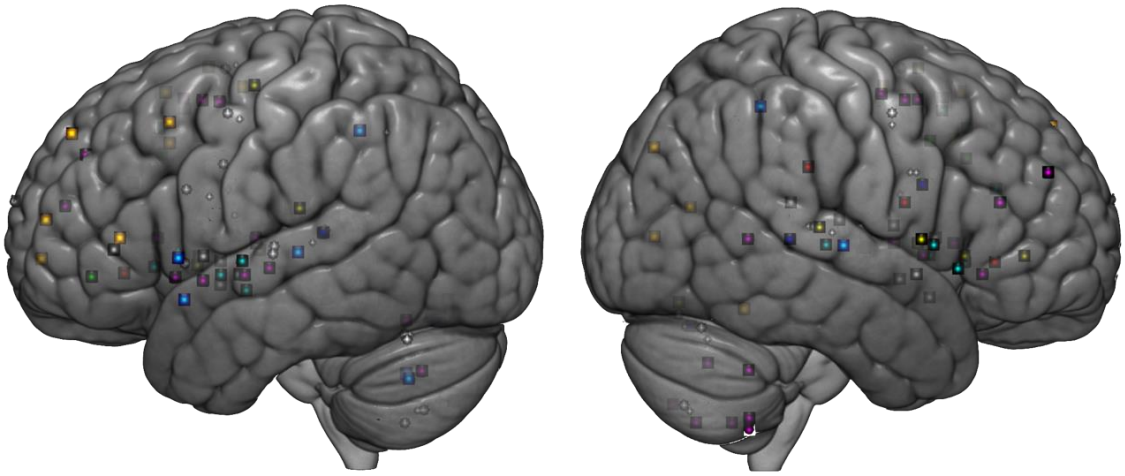
13.4 Neural correlates of rhythmic syntactic processing: An ALE meta-analysis

Beat-based encoding of rhythm is largely implemented in a timing network (for review, see Merchant, Harrington, & Meck, 2013). This timing network comprises two functionally distinct but complementary neural circuits: one is the olivocerebellar-thalamocortical circuit underlying absolute, duration-based timing and the other consists of the CBGT circuits involved in relative, beat-based timing (Teki, Grube, & Griffiths, 2012; Teki, Grube, Kumar, & Griffiths, 2011). In a perception experiment using fMRI, Teki et al. (2011) showed that the

former network is activated when the rhythmic context of the preceding intervals are irregular, while the latter is involved in processing regular, predictable intervals resulting from periodic sequences. To achieve accurate timing, both complementary circuits are activated in a coordinated way (Teki et al., 2012).

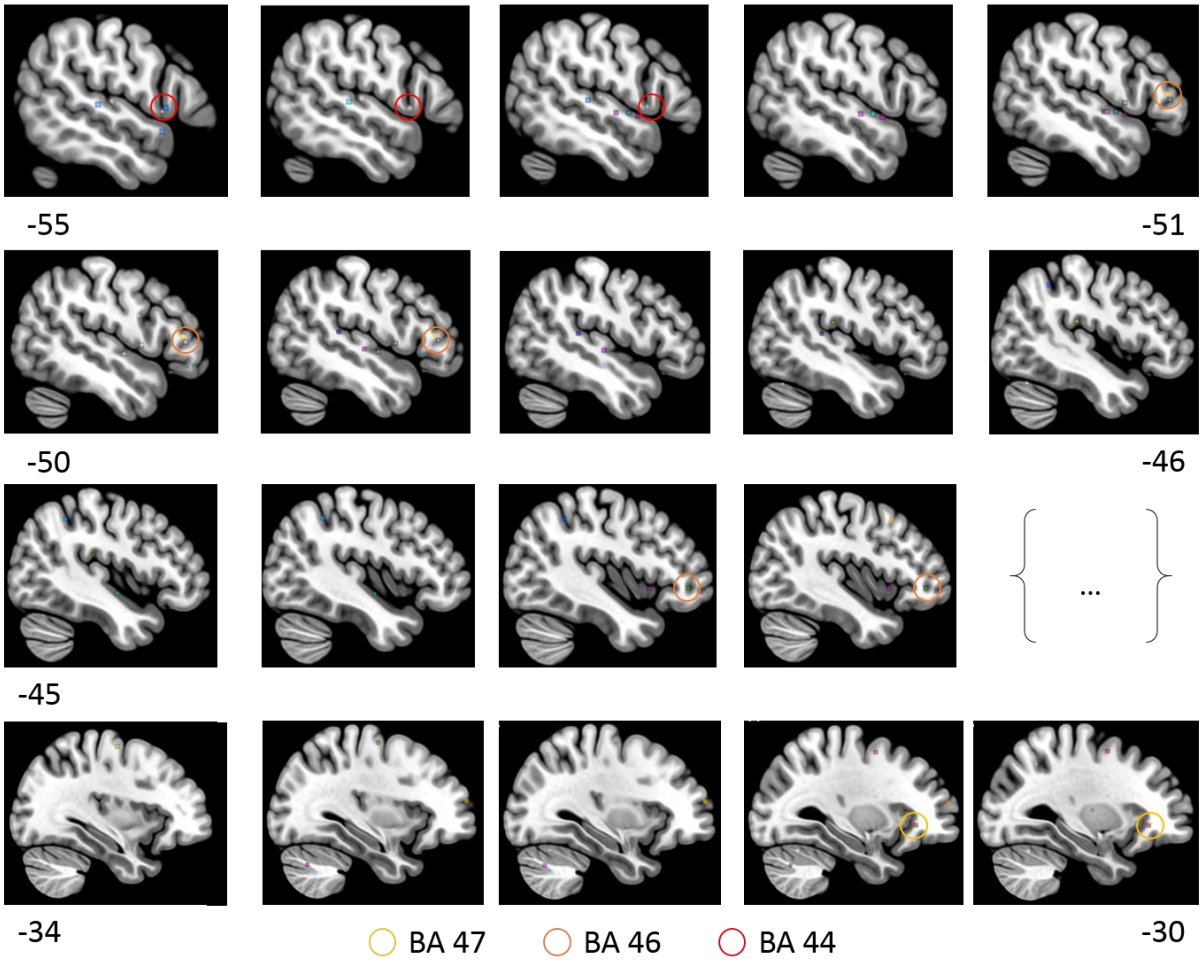
In addition to the timing network, neural correlates of rhythmic syntax might include other regions supporting hierarchical processing, especially areas in the inferior frontal gyrus (IFG) (see Figure 13.6). For example, polyrhythm processing shows bilateral or left BA 47 activation in musicians (Vuust, Roepstorff, Wallentin, Mouridsen, & Østergaard, 2006; Vuust, Wallentin, Mouridsen, Østergaard, & Roepstorff, 2011). Offbeat tapping in non-musicians engages the right BA 47 (Mayville, Jantzen, Fuchs, Steinberg, & Kelso, 2002) and the right BA 44 (Jantzen, Steinberg, & Kelso, 2002). Beat abstraction involves the right BA 44/6 and the right BA 47 (Kung, Chen, Zatorre, & Penhune, 2013). In addition, listening to musical rhythm activates the left BA 44/6 (Chen, Penhune, & Zatorre, 2008a), the left BA 44 (Konoike, Mikami, & Miyachi, 2012), the left BA 45/46 (Bengtsson et al., 2009), and the left BA 46 (Grahn & Brett, 2007) in non-musicians.

a)



b)

Left hemisphere



Right hemisphere

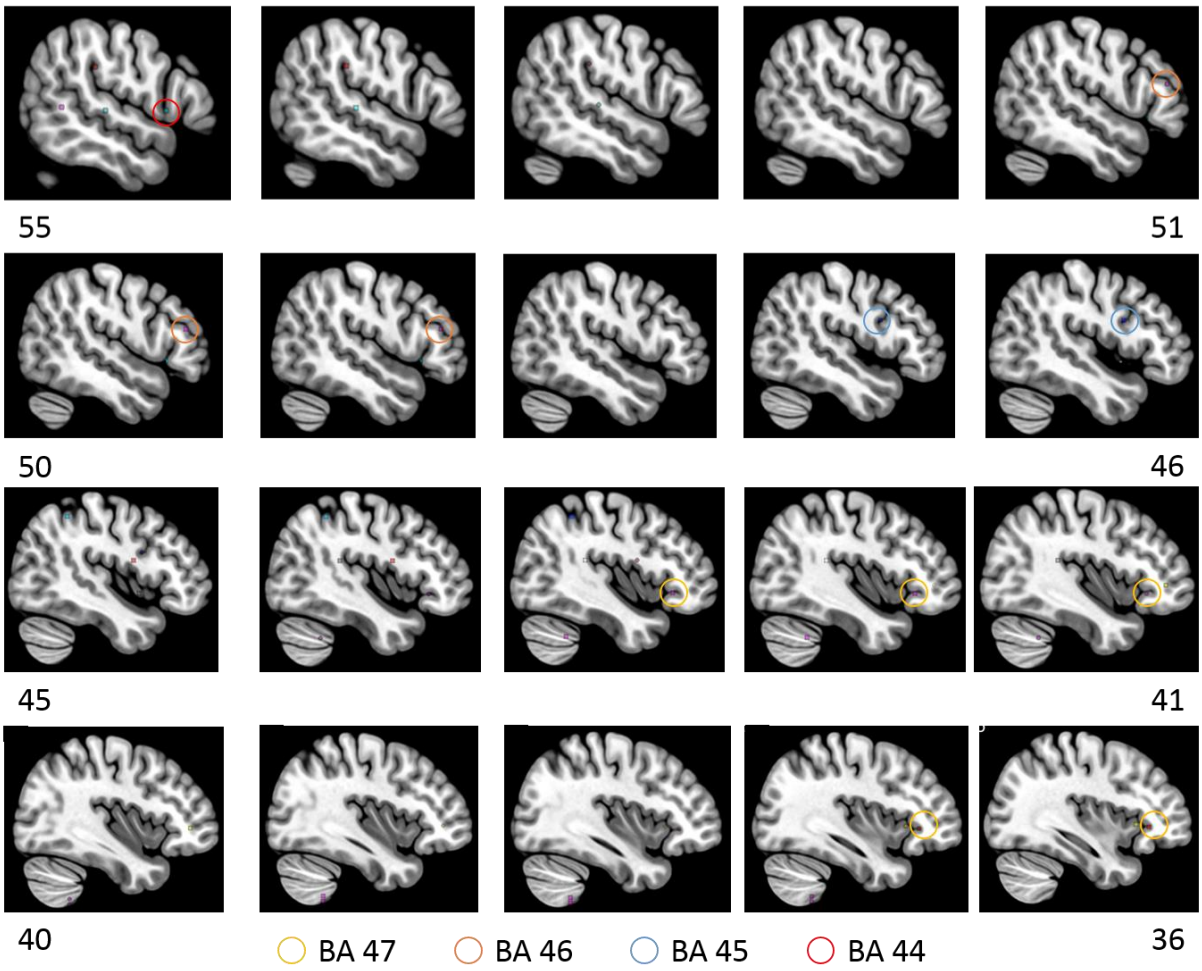


Figure 13.6 Activation peaks in the inferior frontal gyrus (IFG) in rhythmic syntactic processing. The figure is an overview of the peak activation foci (in MNI space) reported in the above-mentioned studies. Coordinates reported in Talairach space were transformed by using convert foci function (Talairach to MNI) implemented in the BrainMap GingerALE software. If a paper mentioned SPM or FSL, the options "Talairach to MNI (SPM)" or "Talairach to MNI (FSL)" were used. The figures were created by MRIcroGL software. a) Lateral view: Peak activations are projected onto the nearest surface. The intensity/brightness of the color represents the distance of the foci to the surface. b) Sagittal view: Peak activation foci in Broca's region and the temporal lobe are labeled with Brodmann areas.

To identify regions constantly involved in rhythmic syntactic processing, an ALE meta-analysis was conducted. First of all, neuroimaging studies were retrieved by using PubMed with "syncopation fMRI", "music rhythm fMRI", and "music beat fMRI" (Date: December 29, 2018). The experiments not using fMRI and using only non-metrical stimuli were excluded.

Further, only one of the experiments that measured exactly same participant group was included. As the number of perception studies were limited, both perception and production studies were included. The experiments investigating both tonal and rhythmic processes were also excluded. Moreover, the experiments that tested listening to or tapping in synchrony with isochronous sequence were excluded as they could mask other rhythmic syntactic regions because of a large number of experiments. In addition, they were analyzed in other meta-analyses and reviews (Chauvigné et al., 2014; Grahn, 2012; Leow & Grahn, 2014; Repp & Su, 2013; Wiener, Matell, & Coslett, 2011; Wiener, Turkeltaub, & Coslett, 2010) so that the results can be still discussed in relation to the current meta-analysis. There was no study testing affect generated solely by rhythm. The 18 experiments included in the current meta-analysis are listed in the *Table 13.1*.

Table 13.1 List of the experiments which entered the ALE meta-analysis on rhythmic syntactic processing. Only the first author's name of a study is displayed in the table.

Rhythmic syntactic processing		Experiments	
Beat abstraction		Chen (2008a) Ex1 (N = 12)	Anticipatory listening > Silence
		Chen (2008a) Ex2 (N = 12)	Anticipatory listening > Silence
		Grahn (2009) (N = 36)	Beat-based rhythm > Non-beat rhythm
		Kung (2013) (N = 11)	Beat finding > Listen isochronous cues
Processing structural relationship	Meter and Grouping	Konoike (2012) (N = 17)	Rhythm encoding > Number encoding
		Grahn (2007) (N = 27)	Metrically simple > Metrically complex & non-metric
		Bengtsson (2009) (N = 17)	Rhythmic > Random
	Syncopation	Mayville (2002) (N = 9)	Syncopation > Synchronization
		Jantzen (2002) (N = 8)	Syncopation > Rest

		Jantzen (2005) (N = 12)	Syncopation > Synchronization
		Jantzen (2007) (N = 9)	Syncopation > Synchronization
		Oullier (2005) (N = 15)	Imagine syncopation > Imagine synchronization
		Herdener (2014) (N = 22)	Main effect of syncopation
		Chen (2008b) (N = 12)	Covariation with complexity
		Chapin (2010) (N = 13)	Attend to auditory rhythm (phase 2) > Rest
Processing structural ambiguity	Polyrhythm	Vuust (2006) (N = 18)	Tap main meter (M) to counter meter (C) > Tap M to M
		Thaut (2008) (N = 12)	Polyrhythmic tapping > Listening
	Drum break	Danielsen (2014) (N = 19)	Transition > Continuous
N of experiments = 18, N of subjects = 281, N of foci = 213			

The ALE meta-analysis was carried out using BrainMap software GingerALE version 2.3.6 (http://www.brainmap.org). First, the foci reported in the Talairach space were converted into MNI space by using convert foci function (Talairach to MNI) implemented in the GingerALE. If the papers mention SPM or FSL, the options "Talairach to MNI (SPM)" or "Talairach to MNI (FSL)" were used. Second, the ALE meta-analysis was performed. This is a coordinate-based analysis of the eighteen experiments which is conducted by using the ALE algorithm as implemented in GingerALE to identify the convergent foci over different studies (Eickhoff et al., 2009; Laird et al., 2005; Turkeltaub et al., 2002). The maps were thresholded by using a cluster-level family-wise error (cFWE) correction ($P < 0.05$) with a cluster-forming threshold of $P < 0.001$ using 1,000 permutations. The results of the analysis are reported in the *Table 13.2*. The anatomical labels were automatically generated by Talairach daemon software (Lancaster et al., 1997, 2000) included in the GingerALE. Third, the ALE image was overlaid

onto the MNI template (Colin27_T1_seg_MNI.nii) by using Mango. The reported clusters and the anatomical labels of the peaks are represented in the Figure 13.7.

Table 13.2 Results of the ALE meta-analysis on rhythmic syntactic processing.

Cluster	BA	MNI coordinates (mm)			ALE (x 10^{-2})	Cluster size (mm^3)
		x	y	z		
1 (Left)	6	-2	-4	58	1.97	4144
	6	2	16	54	1.51	
	6	2	-4	70	1.28	
2 (Left)	*	-30	-64	-26	2.21	2216
3 (Left)	22	-54	8	0	1.76	1896
	22	-52	-8	-2	1.60	
4 (Right)	*	24	10	6	1.96	1368

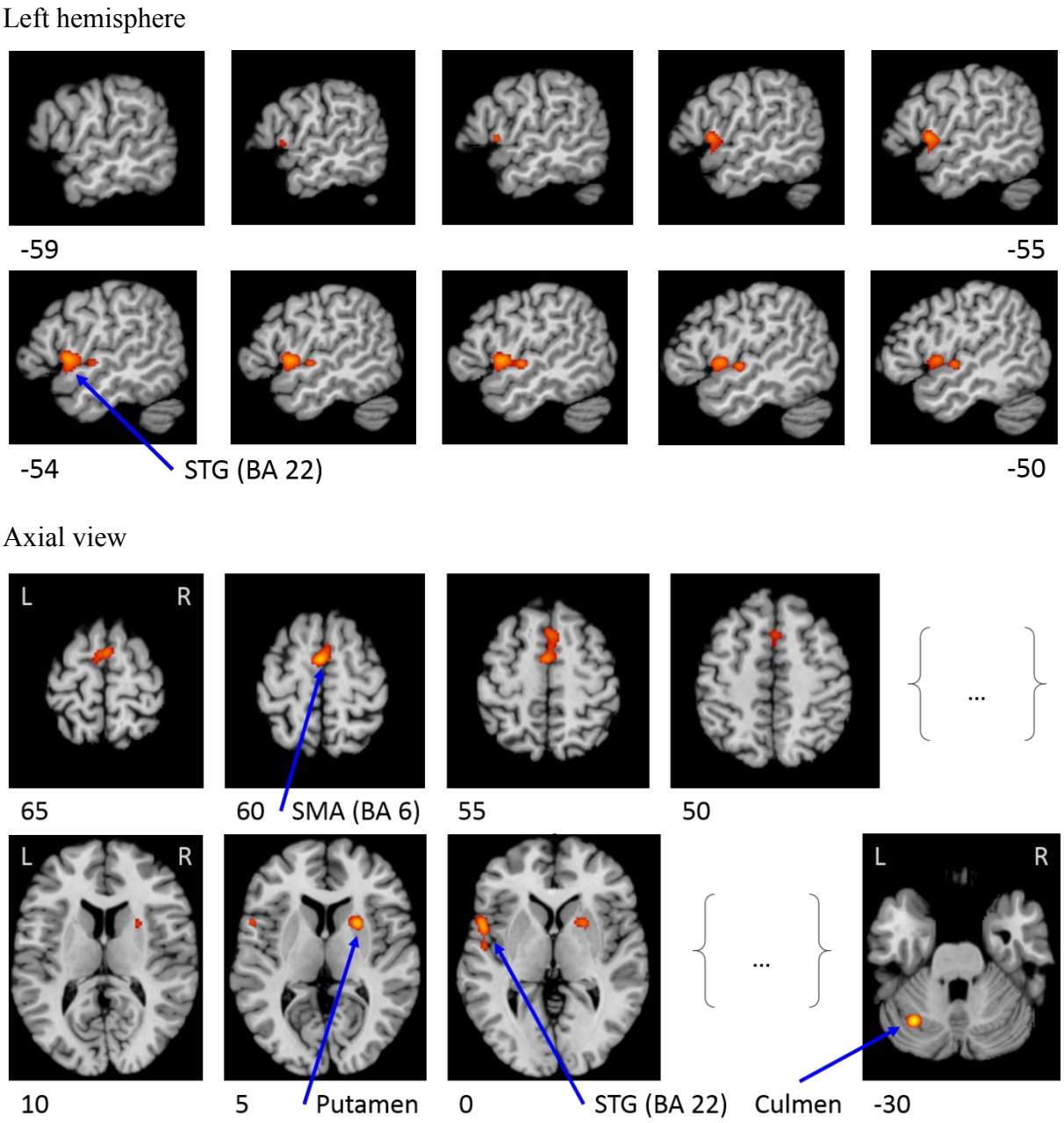


Figure 13.7 ALE images of the meta-analysis on tonal-harmonic syntactic processing overlaid onto the MNI template (Colin27_T1_seg_MNI.nii).

The analysis revealed four significant clusters, i.e., reliable activations over different studies (*Table 13.2*, Figure 13.7, and Appendix A, p. 228). The first cluster consists of the supplementary motor area (SMA; BA 6). The studies investigating beat abstraction (Chen et al., 2008a; Kung et al., 2013), meter and grouping (Bengtsson et al., 2009), syncopation (Chapin et al., 2010; Chen, Penhune, & Zatorre, 2008b; Jantzen, Oullier, Marshall, Steinberg, & Kelso, 2007; Jantzen et al., 2002; Jantzen, Steinberg, & Kelso, 2005; Mayville et al., 2002; Oullier, Jantzen, Steinberg, & Kelso, 2005), and polyrhythm (Thaut, Demartin, & Sanes, 2008)

contributed to this cluster. The second cluster is located in the left cerebellum and peaks in culmen. Studies investigating beat abstraction (Chen et al., 2008a), meter and grouping (Konoike et al., 2012), and syncopation (Chen et al., 2008b; Jantzen et al., 2007, 2005; Mayville et al., 2002; Oullier et al., 2005) contributed to this second cluster. The third cluster is a left hemispheric cluster with peaks in the anterior superior temporal gyrus (aSTG; BA 22) and extensions to the left IFG (BA 44) as well as the left insula (BA 13). Studies investigating beat abstraction (Chen et al., 2008a; Kung et al., 2013), meter and grouping (Grahn & Brett, 2007; Konoike et al., 2012), syncopation (Mayville et al., 2002), and polyrhythm (Thaut et al., 2008) contributed to this cluster. Finally, the fourth cluster is placed in the right basal ganglia, peaks in the right putamen, and extends to the right caudate body as well as globus pallidus. The studies that contributed to this cluster include those examining beat abstraction (Chen et al., 2008a; Grahn & Rowe, 2009; Kung et al., 2013), meter and grouping (Grahn & Brett, 2007), and syncopation (Jantzen et al., 2002; Mayville et al., 2002).

The results show the contribution of the timing network including the SMA, basal ganglia, and cerebellum together with the aSTG to rhythmic syntactic processing. Those regions largely correspond to a subset of the conjunction areas for motor and perceptual sub-second timing studies revealed by an ALE meta-analysis (Wiener et al., 2010). One relevant difference is that Wiener and colleagues (2010) found a peak in the bilateral IFG, while the current meta-analysis in the left aSTG (BA 22) with extensions to the left IFG (BA 44) and insula (BA 13). The most significant difference between the meta-analysis conducted by Wiener and colleagues (Wiener et al., 2010) and the current one lies in the selection of the studies. While the former included all timing related studies regardless or modality and stimulus property, the latter only investigated musical rhythm. The studies examined in the current meta-analysis, therefore, used auditory stimuli with cyclic regularity, i.e., high predictability, possibly leading to a substantial contribution of the left aSTG. Moreover, the third cluster resembles the second cluster in the meta-analysis on the tonal-harmonic syntax (see PART II) and thus is a good candidate for a core region of music syntax. As the restrictions of the current meta-analysis are similar to those discussed in the PART II, I only add a comment that 8 out of 18 studies included in the current meta-analysis investigated syncopation and 5 out of syncopation studies regarded syncopation as off-beat / anti-phase tapping. This type of syncopation only sparsely taps the syntactic aspects of syncopation discussed in the Section 13.1 (p. 118).

In sum, the core neural correlates of rhythmic syntax are implemented in the left aSTG extending to the left IFG (BA 44) and the left insula (BA 13) and a timing network consisting

of two complementary neural circuits including the SMA, basal ganglia, and cerebellum. This indicates an interplay between cortico-subcortical circuits and the temporo-insular pathway for rhythmic syntactic processing. That is, rhythmic syntactic processing has a lot to do with sensory-motor systems in the brain.

14 Neurocognitive mechanisms of rhythmic syntactic processing

14.1 The basal ganglia (BG) and the cortico-basal ganglia-thalamocortical (CBGT) circuits

The basal ganglia (BG) are a set of interconnected subcortical nuclei, located outside of the thalamus and make up the largest subcortical structures in the human forebrain (see Figure 14.1 for a detailed anatomical description). The striatum, consisting of the putamen and caudate nucleus, primarily (95%) consists of medium spiny neurons (MSNs) (Graybiel, 2000; Matell & Meck, 2004). It largely receives cortical inputs beside inputs from the thalamus and the midbrain and projects to the other BG structures (Matell & Meck, 2004). By means of their output nuclei such as the internal segment of globus pallidus and substantia nigra pars reticulata, the basal ganglia nuclei project to thalamus, which in turn projects to different cortical areas (Graybiel, 2000). Thus, together with several cortical areas and the thalamus, the basal ganglia nuclei are organized into circuits.

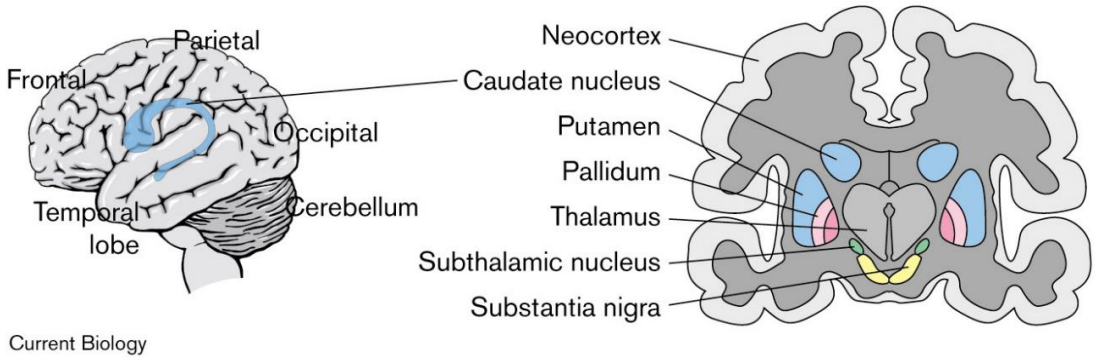


Figure 14.1 The basic anatomy of the basal ganglia. The putamen and caudate nucleus (also caudate) make up the striatum, which receives the most of the cortical input. The pallidum (also globus pallidus; GP) consists of internal and external segments (also globus pallidus internal and external; GPi and GPe) and receives most of the output of the striatum. The subthalamic nucleus (STN) is a key structure controlling pallidal function by receiving inputs from GPe and projecting to GPi. The substantia nigra (SN) entails a dopamine-containing region, pars compacta (SNc), and another region, pars reticulata (SNr). The figure is adopted from Graybiel (2000, p. R509), Current Biology, 10 (14): R509-R511, with permission by Elsevier.

Those circuits involve at least[58] five distinctive networks mainly studied in non-human primates' brains: the "motor", "oculomotor", "dorsolateral prefrontal", "lateral orbitofrontal", and "anterior cingulate" circuits (Alexander, Delong, & Strick, 1986). Together, they are called "cortico-basal ganglia-thalamocortical circuits" (CBGT circuits). Recurrent circuits are also called "loops". Within each circuit, corticostriate inputs are progressively integrated and sent back to a single cortical area. The 'funneling' of the inputs are carried out gradually at striatal, pallidal / nigral, and thalamic levels (see Figure 14.2). Importantly, such funnelling occurs largely in a very 'modular' way, i.e. only *within* the segregated functional circuits (Alexander et al., 1986), thus called "parallel circuits" (Haber, 2003, 2016). Remarkably, however, there are also mechanisms (so-called "integrative circuits") enabling information flow through those functionally segregated circuits (Haber, 2003, 2016).

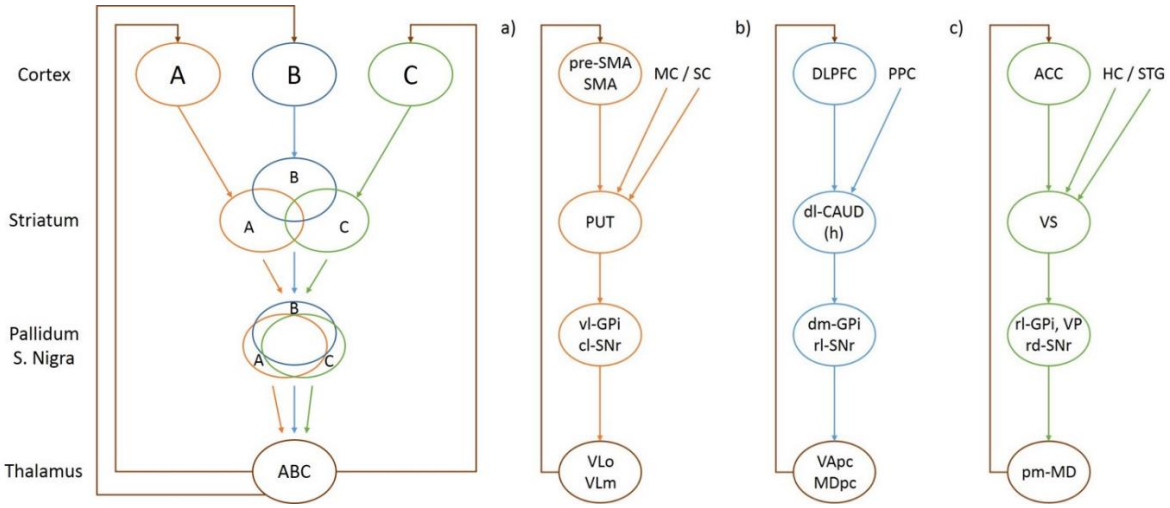


Figure 14.2 Parallel circuits. Different CBGT circuits run in parallel, i.e. projection targets of each circuit differ from each other. The inputs are 'funnelled' gradually at striatal, pallidal / nigral, and thalamic levels. Circuits A, B, and C are corresponding to the circuits a) motor circuit, b) dorsolateral prefrontal circuit, and c) anterior cingulate circuit, respectively. Abbreviations are as follows: ACC (anterior cingulate cortex); CAUD (caudate), (h) (head); GPi (globus pallidus internal); HC (hippocampal cortex); MC (motor cortex); MD (medialis dorsalis); MDpc (medialis dorsalis pars pavocellularis); pre-SMA (pre-supplementary motor area); PPC (posterior parietal cortex); PUT (putamen); SC (somatosensory cortex); SMA (supplementary motor area); SNr (substantia nigra pars reticulata); STG (superior temporal gyrus); VApc (ventralis anterior pars parvocellularis); VLm (ventralis lateralis pars medialis);

[58] Recently, much more details of those circuits have come to be identified. To discuss and describe all those circuits is beyond the scope of the current thesis. For further details, see for example Haber (2003, 2016), Lehéricy, Ducros, Krainik, et al. (2004), Lehéricy, Ducros, Van De Moortele, et al. (2004), Middleton & Strick (2000).

VLo (ventralis lateralis pars oralis); VP (ventral pallidum); VS (ventral striatum); cl- (caudolateral); dl- (dorsolateral); dm- (dorsomedial); pm- (posteromedial); rd- (rostrodorsal); rl- (rostrolateral); vl- (ventrolateral). The figure is based on Alexander et al. (1986).

The function of each circuit is determined on the basis of the cortical areas providing input to the basal ganglia and also the part of the striatum receiving this input. For example, within the "motor" circuit (Figure 14.2a), the cortical projections from supplementary motor area (SMA) and other cortical sensory-motor areas mainly[59] terminate in putamen, while within the "dorsolateral prefrontal" circuit (Figure 14.2b), cortical projection terminates largely[60] within different striatal parts, namely parts of caudate nucleus. The (dorsolateral) prefrontal circuit is also called "executive" or "cognitive" circuit (Watkins & Jenkinson, 2016). Within the "anterior cingulate" circuit (Figure 14.2c), projections from e.g. the limbic structures (such as the hippocampus (HC) and the amygdala) and the anterior cingulate cortex terminate in ventral striatum including nucleus accumbens. The further projection chains via globus pallidus (GP), substantia nigra (SN), and thalamic nuclei back to the cortical area also differ from each other within those distinctive circuits (see Figure 14.2a, b, and c respectively; see also Alexander et al., 1986; Haber, 2003, 2016; Lehéricy, Ducros, Krainik, et al., 2004; Lehéricy, Ducros, Van De Moortele, et al., 2004; Middleton & Strick, 2000 for details).

Concerning the cortical projections toward the basal ganglia, it is worth noting that those projections build a "functional gradient of inputs" within the striatum. For example, diffusion tensor imaging (DTI) studies with human subjects showed that the tracts projecting from the motor cortex and SMA are mainly directed to the posterior putamen (sensorimotor compartment of the putamen) whereas pre-SMA projections are directed to more rostral parts of the striatum (Lehéricy, Ducros, Krainik, et al., 2004; Lehéricy, Ducros, Van De Moortele, et al., 2004). The fibres originating from the prefrontal cortex end up in the anterior striatum (associative compartment of the striatum) and the head of the caudate nucleus (Lehéricy, Ducros, Van De Moortele, et al., 2004). Further, the medial and orbital prefrontal cortex terminates in the ventromedial part, the dorsolateral prefrontal cortex in the central area, and the motor cortex in the dorsolateral region (Haber, 2003). Those results are suggested to be in

[59] There are also 'bridging' projections from the rostral premotor areas to both the caudate nucleus and putamen (Haber, 2003).

[60] Cf. "The DLPFC projects most densely to the rostral striatum including both the caudate n. and the putamen, rostral to the anterior commissure. [...] While there are few terminals in the central and caudal putamen posterior to the anterior commissure, the caudate n. does remain innervated" (Haber, 2003, p. 320).

line with findings from several monkey studies (Haber, 2003; Lehéricy, Ducros, Krainik, et al., 2004; Lehéricy, Ducros, Van De Moortele, et al., 2004).

Moreover, there are two distinctive pathways going through the basal ganglia known as the cortico-striatal and cortico-subthalamo-pallidal pathways (Figure 14.3; for details, see e.g. Graybiel, 2000; Graybiel & Mink, 2009; Matell & Meck, 2000, 2004; Watkins & Jenkinson, 2016). Within the cortico-striatal pathways, further distinctions are made in terms of the direct pathway and indirect pathway. Metaphorically, the direct pathway, leading to decreased inhibition (i.e. release) of the thalamus to excite the cortex, parallels the accelerator of a car sending a 'go' signal, while the indirect pathway, leading to increased inhibition of the thalamus and thus decreased activity in the cortex, can be seen as the brake sending a 'no-go' signal. Within the cortico-subthalamo-pallidal pathway (also hyperdirect pathway), cortical projections 'skip' the striatum and terminate in STN, leading to the rapid inhibition of basal ganglia output. Besides those cortical inputs, the striatum gets dopaminergic inputs from the SNc, which most commonly co-occur with glutamergic inputs. In addition, the striosomes (striatal bodies) are interconnected with the SNc and regulate its functions (Figure 14.3).

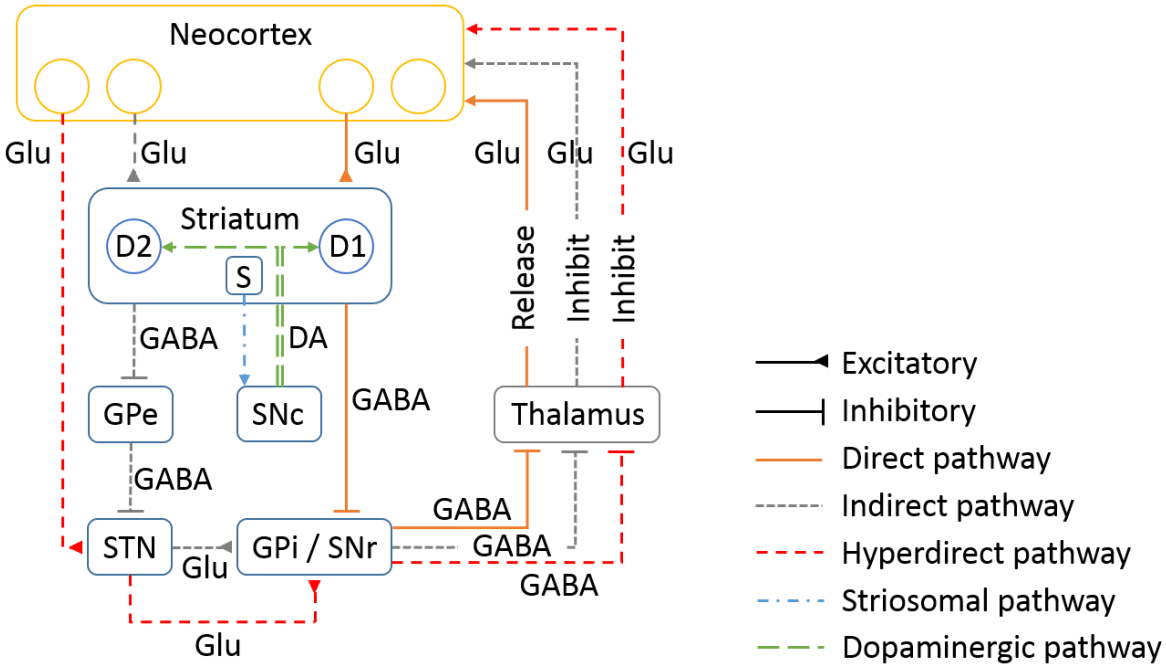


Figure 14.3 Cortico-striatal and cortico-subthalamo-pallidal pathways. The CBGT circuits include different pathways: direct pathway, indirect pathway, and hyperdirect pathway. The direct pathway disinhibits the thalamus, leading to excitation of the cortex. Indirect and hyperdirect pathway inhibit the thalamus, i.e. resulting in decreasing cortical activities. Abbreviations are as follows: D (dopamine receptor); DA (doperminergic); GABA

(GABAergic); Glu (glutamergic); GPe (globus pallidus external); GPi (globus pallidus internal); S (striosome); SNc (substantia nigra pars compacta); SNr (substantia nigra pars reticulate); STN (subthalamic nucleus).

A balanced combination of these pathways is very important for the normal function of the basal ganglia. Dysfunctions at that level cause several motor and cognitive impairments. For example, Parkinson's and Huntington's disease are neurodegenerative diseases and belong to classical basal ganglia disorders. Parkinson's disease (PD) is caused by degeneration of the dopamine-containing neurons in the SNc, which in turn disables to supply the striatum with dopamine (Graybiel, 2000; Watkins & Jenkinson, 2016). The loss of dopamine innervation in the striatum first affects the putamen, resulting in poor motor output. Huntington's disease (HD) is caused by degeneration of GABAergic neurons in the striatum (MSNs in the indirect pathway), resulting in movement control deficits (e.g. choreiform or "dancing") (Graybiel, 2000; Watkins & Jenkinson, 2016).

The roles of the striatal MSNs and the dopaminergic inputs from the SNc have been discussed extensively by several authors (e.g., Coull, Cheng, & Meck, 2011; Graybiel & Mink, 2009; Matell & Meck, 2004; the description below refers mainly to the last authors). Each striatal MSN receives 10,000 to 30,000 separate inputs from cortical and thalamic neurons that are considered to represent particular tasks and abstract rules by firing in an oscillatory fashion at a particular rate. The coherent number of coincidental cortical or thalamic activations provide simultaneous inputs to the striatum which are necessary for the striatal MSNs to fire. Thus, the striatal MSNs act as a coincidence detector. Moreover, the firing threshold of the striatal MSNs is considered to be modulated by the GABAergic inhibitory striatal interneurons and the non-burst, tonic dopaminergic input from SNc. Burst-mode, phasic dopaminergic input from SNc is suggested to modulate the direction of synaptic strength change, cause the long-term potentiation or depression, and thus serve as the reinforcement signal.

Though the role of the BG in motor control (which is the most well-known and uncontroversial function of the basal ganglia) has long been emphasized in research, it is now widely accepted that the BG, as parts of the CBGT circuits, function in a wide range of cognitive processes at this circuit level (Graybiel, 2000; Graybiel & Mink, 2009; Lieberman, 2002, 2010, 2016; Middleton & Strick, 2000). For example, the basal ganglia have been claimed to be involved in sequence learning and control (Ullman, 2006; Ullman & Pierpont, 2005), cognitive set-shifting (Hochstadt, Nakano, Lieberman, & Friedman, 2006; Monchi, Petrides, Petre, Worsley, & Dagher, 2001), attention (Grossman, 1999), and working memory

(Harrington, Zimbelman, Hinton, & Rao, 2010; Hochstadt et al., 2006; Lustig, Matell, & Meck, 2005). The BG and the CBGT circuits are also repeatedly suggested to play a crucial role in processing linguistic syntax (e.g., Dominey & Inui, 2009; Friederici & Kotz, 2003; Lieberman, 2002; Ullman, 2006) and musical rhythm (e.g., Grahn, 2009; Leow & Grahn, 2014). In the following subsections, I review the studies investigating the role of the BG and the CBGT circuits in linguistic syntax and beat-based encoding of rhythm.

14.2 Cognitive and neural processes implemented in the CBGT circuits

As noted above, any given temporal sequence is ambiguous and the interpretation of a rhythmic sequence changes depending on which underlying hierarchical representation the listener may choose. Thus, how people extract primary beats from musical surface is based on which hierarchical structure is represented or built up online in the listener's mind/brain. The ambiguity in terms of multiple conflicting representations was best posed in the examples of polyrhythm and hemiola. In both cases, a flexible switching between conflicting representations is required. That is, a parser should keep conflicting representations in parallel and this is in favor of a variable choice model of parsing where the parser selects a final analysis on the basis of a competition between alternative analyses, i.e., a winner-take-all principle. Concerning musical rhythm, switching between well-formed, yet conflicting representations was suggested to figure-ground switching (London, 2012b; Vuust & Witek, 2014). As Jackendoff (1991, p. 214) puts it, "one can attend to only one analysis at a time". The parser evaluates plausibility of various representations by *integrating* multiple sources of information and then selects the currently most salient interpretation. This process is implemented by a *selection* function (Jackendoff, 1991). Some contextual information, e.g., dance or song in case of African drum ensemble music, *weight* one interpretation over another. Moreover, keeping multiple representations in parallel requires a *maintenance* function.

The cognitive processes such as integration, selection, weighting, and maintenance have an appropriate granularity so that they can be implemented by neural processes. In the following, I argue that the CBGT circuits belong to neural mechanisms that implement those cognitive processes. First, as reviewed above, the MSNs in the striatum serve as integrators by detecting coincidental inputs from cortical areas. Second, by inhibiting or releasing cortical representations, the basal ganglia are able to select a salient representation. The selection function of the basal ganglia is carried out based on the winner-take-all principle (Graybiel & Mink, 2009). Third, the dopaminergic input from the SNc into the striatum weights cortical inputs differently by modulating the firing thresholds of the MNSs. The representations are

maintained in the (pre)frontal cortex by recurrent excitation, with the aid of the basal ganglia functions if conflicting representations should be maintained. These interpretations are in line with computational and functional models of the CBGT circuits developed in research on working memory (Frank, Loughry, & O'Reilly, 2001; Lustig et al., 2005), functional models of the frontal cortex (Fuster, 2008a; E. K. Miller & Cohen, 2001), and a general functional architecture implemented in the ACT-R (Anderson et al., 2004). Syntax as mapping between hierarchical structure and temporal sequence can be implemented based on those cognitive and neural processes.

Before proceeding in the next section with a discussion of how those cognitive and neural processes lead to rhythmic syntactic processing, I would like to point out further relevant aspects. Concerning the representations in the prefrontal cortex, as already mentioned in PART II, there is a rostro-caudal gradient of memory, control, and goal representation in the frontal cortex with motor part in most of the caudal part and cognitive or abstract part in the most rostral part (Badre & D'Esposito, 2009; Fuster, 2008b; Koechlin & Jubault, 2006; Uddén & Bahlmann, 2012). As mentioned in the Section 14.1 (p. 137), different cortical areas preferably project to different parts of the striatum, with motor-related areas projecting to the putamen and more cognitive or associative areas to the caudate. Therefore, it can be hypothesized that the same above-mentioned cognitive and neural processes account for syntax in both motor and cognitive domains, with the former relying more heavily on the motor circuit and the latter on the cognitive circuit. Thus, different configurations of the CBGT circuits account for different cognitive systems such as language and music.

One promising avenue may be cognitive control (E. K. Miller, 2000; E. K. Miller & Cohen, 2001), i.e. control of lower-level sensory, memory and/or motor operations to achieve a goal (internally represented). For instance, E. K. Miller and Cohen (2001, p. 170) built on "the fundamental principle that processing in the brain is competitive: Different pathways, carrying different source of information, compete for expression in behavior, and the winners are those with the strongest sources of support". This is exactly the principle of neural computation that is consistent with one suggested to be implemented in the BG and the CBGT circuits (Coull et al., 2011; Dominey & Inui, 2009; Graybiel & Mink, 2009; Matell & Meck, 2004; Sambin et al., 2012; Teichmann, Dupoux, Cesaro, & Bachoud-Lévi, 2008). Moreover, the role of the BG and the CBGT circuits in cognitive control has repeatedly been suggested both in theoretical (Lieberman, 2002, 2010, 2016) and experimental research (Monchi et al., 2001), as well as in modelling (Caso & Cooper, 2001).

Jackendoff (1991, p. 226) adds that "[the] tension among the conflicting analyses may also surface in the listener's experience as affect, even though only one of the analyses is experienced at the moment as "the" structure of music". Polyrhythmic tension is a good example and provides strong evidence for a variable choice model of parsing (see PART II, Section 14.2, p. 142) as multiple rhythmic interpretations exist in parallel. By selecting the most salient representation at a given moment, the CBGT circuits contribute to resolve such tension between conflicting analyses. The function of the limbic circuits including the limbic systems and the ventral striatum thereby is a relevant issue, but beyond the scope of the current thesis. This issue together with the interplay between STG, BA 44, and insula in processing affect (Molnar-Szakacs & Overy, 2006) might be a promising future direction. Finally, the cerebellum's role in sensory-motor integration, sensory prediction, and feedback error processing (e.g., Kotz, Brown, & Schwartze, 2016 for discussions) should be also integrated in the current model.

14.3 The CBGT circuits for rhythmic syntactic processing

The current meta-analysis on rhythmic syntactic processing points out the involvement of the CBGT circuits, especially of the motor circuit through the putamen. Importantly, however, different parts of the basal ganglia, especially different striatal components, and the different CBGT circuits seem to be involved in (at least) three distinctive sub-processes: beat finding, beat continuation, and beat adjustment (Grahn & Rowe, 2013). "Beat finding" means abstracting the underlying regular beat and thus corresponds to beat abstraction as introduced above. "Beat continuation" refers to continually generating and predicting the internal beat, and thus parallels to beat generation introduced above. "Beat adjustment" relates to processing prediction errors and updating the subsequent predictions. Those sub-processes might recruit cognitive processes implemented in the CBGT circuits differently or combine them with processes implemented in other networks, thus requiring further differentiations (Leow & Grahn, 2014).

Among those sub-processes, the putamen is most strongly engaged in beat continuation, i.e. during continually predicting the timing of future events at the same rate without tempo change (Grahn & Rowe, 2013). The parallel activation of the SMA during beat continuation in this experiment indicates that the motor CBGT circuit is involved in beat continuation. Thus, the motor CBGT circuits involving the SMA and the putamen seem to play a particularly important role in maintaining periodic predictions (Leow & Grahn, 2014). A single-cell recording study with two rhesus macaques showed that the cells in their medial premotor cortex

(corresponding to the SMA) are tuned to interval durations ranging from 450 ms to 1000 ms during the synchronization-continuation task and a single interval reproduction task (Merchant et al., 2013), indicating that the SMA represents different time intervals. During the continuation phase, the SMA and the putamen selecting and maintaining the particular time interval over others could be more strongly engaged, as there is no external cues that synchronize cortical cell firing and provide simultaneous inputs to the MSNs.

The caudate, in contrast, seems to play a more important role in beat finding, i.e. generating the beat in concert with the external cues, in a perceptual task (Kung et al., 2013). In this experiment, the caudate was engaged together with several cortical regions such as the STG and the VLPFC whose activity was sensitive to metrical complexity or beat strength.[61] VLPFC showed increasing connectivity with the right DLPFC and bilateral BG at the caudate-putamen border in correspondence with increasing metrical complexity. Moreover, based on some studies outside of the research on music, the caudate was suggested to be involved in processing prediction errors (Grahn & Rowe, 2013). For example, the head of caudate nucleus was suggested to be engaged especially in processing breaches of expectation during judgement task in which participants indicated whether the presented dance performance was correct, based on the rules they previously learnt (Schiffer & Schubotz, 2011). In beat finding, the integration of auditory and motor cues is required to generate the beat (Kung et al., 2013) and prediction errors should be processed till the beat got generated and automated. Thus, the functions of the caudate in beat-based encoding of rhythm can be hypothesized as generating the beat in concert with external cues by integrating them and processing prediction errors, possibly with a strong collaboration with the cerebellum.

Moreover, PD patients show impaired beat-based encoding of rhythm. While their general perceptual and motor timing deficits were claimed (C. R. G. Jones & Jahanshahi, 2014), the studies which I reviewed show impairments rather in particular aspects of timing abilities. For example, timing deficits of PD patients are especially pronounced in perceptual tasks requiring maintenance of internal periodic prediction to discriminate durational cues or rhythms (Benoit et al., 2014; Cameron et al., 2016; Grahn & Brett, 2009). Other perceptual experiments in which physical cues seem to have provided enough information to succeed in the tasks,[62] in contrast, showed intact beat-based timing in PD (Cameron et al., 2016; Geiser &

[61] A stimulus is metrically complex or the beat strength of a stimulus is weak, if external cues don't occur coincidentally with the underlying periodic beat.

[62] The stimulus material used by Geiser and Kaelin-Lang (2011) contained very simple repetitive patterns, in which strong beat positions (i.e. 'one' and 'three' in 4/4 meter) always underwent physical cues. In the beat

Kaelin-Lang, 2011; but see Benoit et al., 2014). Motor timing deficits of PD patients depend very much on metronome tempo[63] (Benoit et al., 2014; N. S. Miller et al., 2013) and are more evident in the continuation phase where the external cue disappears and internal periodic prediction should be maintained, than in the synchronization phase where one taps along the physical cues (Benoit et al., 2014; Tolleson et al., 2015). Moreover, variability in timing tasks was especially pronounced only in a subset of the PD group (Merchant, Luciana, Hooper, Majestic, & Tuite, 2008). To my knowledge, there is no study on beat-based encoding of rhythm conducted with HD patients in a comparable way, but some studies showed deficits of general timing ability in HD patients (Cope, Grube, Singh, Burn, & Griffiths, 2014; Thaut, Miltner, Lange, Hurt, & Hoemberg, 1999).

In sum, the motor circuit including the SMA and the putamen plays a central role in beat-based encoding of rhythm, especially in continually predicting the internal beat. Moreover, the executive circuit including the VLPFC and the caudate preferably involved in beat finding, especially when the internal beat is hard to generate. Regardless of the limited number of studies, it can be said that the results of patient studies support the hypothesized functions of the BG and the CBGT circuits in beat-based encoding of rhythm, i.e. selecting and maintaining the beat.

15 On the relationship between syntax in language and rhythmic syntax

15.1 Identifying the missing link from theoretical and empirical perspectives

Language-music comparative research mainly focused on metrical structures of music and speech, which are often regarded as similar in terms of hierarchical metrical grids (Fabb & Halle, 2012; Fitch, 2013; Jackendoff, 2009; Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1983). Each note in music and each syllable in speech gets a beat at the lowest level of hierarchy, which is projected onto the higher levels if it is more salient than the others. In this way, metrical grids of music are formally homologous to those of phonology (Jackendoff & Lerdahl, 2006). In addition, Lerdahl and Jackendoff (1983) pointed out parallels between prosodic structure and time-span structure. Prosodic hierarchy consists of syllables grouped into a foot,

alignment tests (BATs), the task can be solved by simply evaluating the alignment of two physical cues, i.e. musical excerpts and the metronome ticks. Given a large difference in the stimulus features and numbers, it is hard to make any well-founded proposals why PD patients' performance was impaired in one experiment (Benoit et al., 2014) and not in the other (Cameron et al., 2016). Especially, concerning the BATs, it is important to control for the beat strength of the stimuli by using the model of, for example, Povel & Essens (1985) or Longuet-Higgins and Lee (1984).

[63] It was suggested that the interval range of 400-600 ms is critical for the PD patients (C. R. G. Jones & Jahanshahi, 2014). The study conducted by N. S. Miller et al., 2013 also showed that PD patients were less accurate in tapping along the metronome with 500 ms IOIs in comparison with those with 1000 ms and 1500 ms IOIs.

feet grouped into a phonological word, and phonological words grouped into phrases. Thus, the authors argued that prosodic and time-span structures are both based on "a segmentation of the surface string into a layered hierarchy" (Lerdahl & Jackendoff, 1983, p. 321). This is highly relevant to the idea introduced in the Section 12.3 (p. 114), to regard one aspect of rhythmic syntax, i.e., cyclicity, as a special form of chunking, with accented elements as heads projecting to higher layer for further computation.

Lerdahl and Jackendoff (1983, p. 218) further distinguish prosodic structure from narrow-sense syntactic structure: While "in [narrow sense; RA] syntactic structure a category may recur inside one of its constituents (for instance, NP inside PP inside NP), this is impossible in the phonological segmentation". They then argue that prosodic structure is thus rather similar to time-span structure than narrow-sense syntactic structure. However, this contrast is not fruitful if we take rhythmic syntax seriously and investigate the relationship between language and music. Contrarily to speech allowing one to one assignment of beats to syllables, in music "a single note can subtend multiple beats, and a beat can be subdivided by multiple notes" (Jackendoff, 2009, p. 199). This unique feature of musical meter allows for structural complexity by recursive embedding of beats into beats, which is regarded by Longuet-Higgins and Lee (1984) as parallel to phrasal structures of narrow-sense syntax. Moreover, as discussed above, two-ness and three-ness of meter are categorical and the binary meter can be embedded into ternary meter that is embedded in binary meter (Figure 15.1).

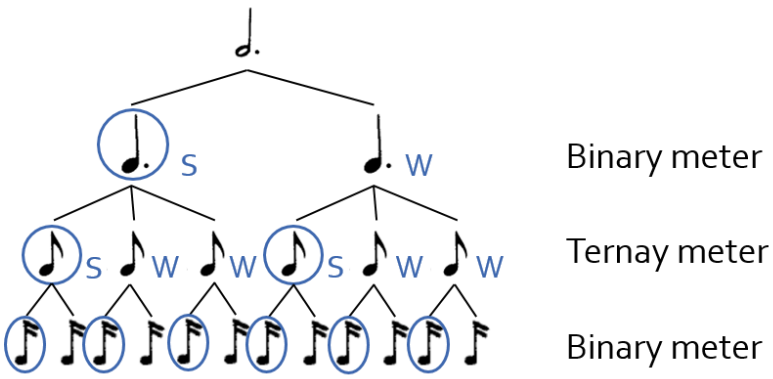


Figure 15.1 Recursive embedding of meter into meter.

In addition, Lerdahl & Jackendoff (1983) do not discuss the relationship between syllable structure and musical rhythm as well as dance because they limit themselves to metrical structure. A syllable consists of an onset, nucleus, and coda. The nucleus is the head of a syllable and thus corresponds to a beat in metrical structure. This structure parallels to that

of action syntax consisting of a preparation, head, and coda as introduced by Jackendoff (2009). In both cases, the hierarchical structure is represented as [[onset/preparation, [nucleus/head, coda]]. In music, ternary meter intuitively is structured in [[strong, weak], weak] and this does not seem to include any preparation element. However, in dancing Waltz (step – turn – feet together), the structure is rather {{step, turn}, feet together} and the feet together is a preparation to the next step, i.e., [feet together, [step, turn]] (Fitch, 2016). A stable body configuration such as two feet together on the ground is often regarded as a preparation to the next unstable motion (Lasher, 1981). Therefore, a person with an experience of dancing waltz might regard the *three* of the waltz music as a preparation to the next *one*, i.e. [weak, [strong, weak]]. This type of structure in phonology, action, and music is not a layered hierarchy, but an asymmetrical hierarchy for action planning.

In sum, rhythmic syntax relates to both narrow-sense syntax and phonological syntax. Of course, rhythmic syntax encodes affect, i.e., tension-relaxation pattern, which has no good parallel in language domain. This is, however, somewhat evident given a difference between language and music in terms of computational problems to be solved. More importantly, rather, narrow-sense syntax, phonological syntax, and rhythmic syntax all have to go with mapping between hierarchical structure (regardless of layered, (recursively) embedded, or asymmetrical planning hierarchy) and temporal sequence.

From empirical perspectives, too, there is an increasing number of studies pointing out a close relationship between linguistic and rhythmic syntactic processing (for a review, see also Gordon, Jacobs, Schuele, & McAuley, 2015). In linguistic sentence comprehension and production, a critical problem for syntactic processing is the mapping between hierarchical predicate-argument structure reflecting *who did what to whom* and surface morphosyntactic and phrasal syntactic configurations of a string (e.g., a sentence) expressed by functional morphemes or word orders (Kaplan, 2003). As for rhythmic syntactic processing, the mapping between hierarchical metrical structures and phenomenal accents should be established to generate the primary beats. Comparative research on language syntactic processing and rhythmic syntactic processing from empirical perspectives largely deals with interaction and transfer effects of linguistic syntactic processing and rhythmic syntactic processing.

For example, several event-related potential (ERP) studies show a tight relationship between rhythm and syntax in auditory language processing. Metric and syntactic violation processes interact in later integration process reflected in P600 (Schmidt-Kassow & Kotz, 2009). Metric regularity also seems to facilitate processing complex sentence structures – the P600 mean amplitude shows reduction in structurally demanding sentences when presented

with a metrically regular speech rhythm (Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013). Behavioral experiments with children, too, demonstrate a strong correlation between rhythm and syntactic processing: 6-year-old children's ability to discriminate rhythmic sequences correlates with their morpho-syntactic processing ability in a language production task (Gordon, Shivers, et al., 2015). Moreover, children with specific language impairment (SLI)[1] show better morphosyntactic processing if an auditory sentence follows a metrically regular cue (Przybylski et al., 2013).

In the following, I argue that shared neurocognitive mechanisms behind the interaction and transfer effects are best investigated by extending research to the subcortical structures. Comparative language-music research on syntax continues to focus on cortical structures, but there is growing evidence that such a discussion could benefit from taking the role of the subcortical structures into account (see, for example, Kotz et al., 2009). Indeed, the CBGT circuits are at the core of rhythmic syntactic processing and thus provide a good starting point for comparative endeavor.

15.2 The CBGT circuits for syntactic processing in language

To date, several models have been suggested to investigate the brain basis of speech as well as language processing and, in discussing syntactic processing, many of them largely focus on cortical structures and cortico-cortical circuits (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2016; Friederici, 2016; Hagoort, 2016; Hickok & Poeppel, 2016). The current review adds to such cortically based models by investigating the role of the BG in syntactic processing and extending the brain circuits under consideration to the CBGT circuits. This extension is of particular importance because there is an increasing number of evidences suggesting involvement and necessity of the BG and the CBGT circuits in language processing. For example, within the language network, the BG, but not other structures, were preferably engaged in representing syntactic constituent structures of language – regardless of written language or sign language (Moreno, Limousin, Dehaene, & Pallier, 2018; Pallier, Devauchelle, & Dehaene, 2011).

The findings especially point out the prefrontal circuit including the ventrolateral prefrontal cortex (VLPFC) and the caudate (so-called "executive" or "cognitive" circuit) in processing linguistic syntax among different parallel CBGT circuits. For example, neuroimaging experiments showed that Broca's area including the BA 44 and BA 45 as well as the caudate are involved in syntactic (i.e. word order) anomaly recognition in pseudoword sentences (Moro et al., 2001) and processing syntactically ambiguous sentences (Stowe et al.,

2004). The caudate activity was not observed during anomaly recognition of phonological rule and agreement in pseudoword sentences (Moro et al., 2001).

Another study further specified the role of the BG in syntax at the network level. By using fiber tracking, voxel-based lesion-symptom mapping (VLSM), and behavioural measure, Teichmann et al. (2015) investigated a genuine role of the Broca-caudate tract in processing canonical and non-canonical sentences (i.e. phrasal syntax) as well as regular and subregular nonce-verbs (i.e. combinatorial morphosyntax). VLSM results suggest that processing noncanonical sentences and subregular nonce-verbs is associated with voxels of a region ranging from the left inferior prefrontal cortex (BA 44, BA 45, rostrally adjacent areas of BA 47), through the intervening white matter, to the head of the left caudate. The largest number of voxels was identified in BA 45. Regular nonce-verb processing was associated with two separated voxel clusters in Broca's region (BA44 and 45) and left caudate head, respectively, and canonical sentence processing with a voxel cluster of BA 44 and 45 extending to white matter, without involvement of deep white matter regions or the caudate.

However, one study reported increased putamen activation together with the caudate in the domain of linguistic syntax (Moreno et al., 2018). Importantly, in this study, the boundaries of constituents appeared very periodically and predictive. Another study reported increased left putamen activation (but no activations in the cortical nodes of the motor circuit) in phrase structure violation processing together with the left frontal operculum and the left anterior superior temporal gyrus (Friederici, Rüschemeyer, Hahne, & Fiebach, 2003). Although the putamen was involved in this study, the motor circuit is unlikely to play a significant role. It is possible that the putamen contributed to execution of the patterns generated in the temporal cortex, which, however, still remains speculative for now.

A series of sentence comprehension as well as judgement studies with explicit and implicit tasks suggests impaired morphosyntactic and phrasal syntactic processing abilities of patients with PD and HD in comparison to control participants. Especially, PD and HD patients show reduced abilities in processing non-overlearned sequences requiring rule-based online manipulation (Bocanegra et al., 2015; Friederici, Kotz, Werheid, Hein, & von Cramon, 2003; García, Bocanegra, et al., 2017; García, Sedeño, et al., 2017; Grossman, 1999; Sambin et al., 2012; Teichmann et al., 2008, 2005). However, their ability to process lexicalized or overlearned morphosyntactic and phrasal syntactic configurations is relatively preserved (Friederici, Kotz, et al., 2003; Longworth, 2005; Teichmann, Dupoux, Kouider, & Bachoud-Lévi, 2006; Teichmann et al., 2015). In addition, there are important between-diseases differences of anomaly patterns.

Concerning PD patients, the difference in canonicity (e.g. active vs. passive voice), in general, seems to have no effect on their performance (Grossman, 1999; Terzi, Papapetropoulos, & Kouvelas, 2005). Rather, presence of alternative interpretation, and thus the need of reordering or altering, seems to affect their sentence comprehension (Bocanegra et al., 2015; Friederici, Kotz, et al., 2003; García, Sedeño, et al., 2017). The PD patients' impairment in controlled syntactic processing seems to stem from limitations in the strategic distribution of cognitive resources such as selective attention (Grossman, 1999). However, the PD patients' impairment does not deal with selective attention in general (Kotz et al., 2009), but especially with internally (and not externally) controlled attention (R. G. Brown & Marsden, 1988).

In support of this hypothesis, BG patients[64] showed normal P300 in an auditory oddball experiment where the deviant is marked by external cue (e.g. by pitch change), but no P600 was displayed in PD patients and the BG patients in syntactic parsing experiments where acceptability of the sentence had to be judged and thus it was required to focus attention according to the internal cue such as predicate-argument structure (Friederici, Kotz, et al., 2003; Frisch, Kotz, von Cramon, & Friederici, 2003; Kotz et al., 2003). Moreover, over-regularization or priming of external rhythmic cue in the stimuli causes recovery of P600 in PD patients because those external cues guide the internal attentional control (Kotz & Gunter, 2015; Kotz & Schmidt-Kassow, 2015). Certain tasks as well as external or surficial morphosyntactic cues also seem to guide the attention of PD patients (Grossman, 1999; Grossman et al., 2003).

Concerning HD patients, their deficits are strongly associated with structural complexity and canonicity of sentences (García, Bocanegra, et al., 2017; Sambin et al., 2012; Teichmann et al., 2008, 2005) because of their more general impairments in suppressing automatically engaged representations (e.g. canonical interpretation) and applying rules to select an alternative representation (e.g. noncanonical interpretation) (Sambin et al., 2012; Teichmann et al., 2008, 2005). Again, in HD patients, too, the selection function is impaired.

Moreover, in relation to the neurophysiological details of the BG and the CBGT circuits reviewed in the Section 14.1 (p. 137), the following hypotheses can be built. First, dopaminergic input to the striatum is considered to have a modulatory and reinforcing effect on striatal MSN activities and 'weights' the cortical (or thalamic) inputs to the striatum, i.e. associates them with a certain relevance and create attentional or 'goal-oriented' biases (Graybiel & Mink, 2009; Haber, 2016; Lustig et al., 2005; Matell & Meck, 2000, 2004; E. K.

[64] Frisch et al. (2003) and Kotz et al. (2003) did not examine PD patients, but patients with BG lesions.

Miller & Buschman, 2008). Thus, abnormal dopaminergic inputs in PD could lead to impaired internal control in associating striatal activities with goals (or meanings), which possibly leads to their difficulties to distribute internal attentional resources between possible alternative interpretations, i.e., malfunctioning of the selection function. Second, degeneration of GABAergic MSNs in HD patients seems to primarily affect inhibitory control to suppress automatic syntactic processes. Moreover, loss of MSNs leads to insufficient integration of cortical inputs encoding different syntactic representations within striatum, which could lead to difficulty in choosing an alternative representation, i.e., again malfunctioning of the selection function.

In sum, the executive circuit, especially the Broca-caudate tract, plays a central role in syntactic processing. Moreover, the BG are preferably involved in processing non-canonical and ambiguous sentences as well as non-overlearned forms. Processing canonical sentences and overlearned forms largely relies on the cortico-cortical activations. The selection, weighting, and integration function implemented in the BG seem to be required for inhibiting automatic structuring and selecting alternative one.

15.3 Shared neurocognitive mechanisms implemented by the CBGT circuits

The results of this chapter reveal a hidden connection between language and music processing. The same neural mechanisms implemented in the same neural structures and circuits, i.e. the BG and the CBGT circuits, are necessary for language and music processing, but degrees of expression on the motor to cognitive continuum within those neural structures and circuits vary between those two domains. The studies reviewed above suggest that the BG and the CBGT circuits are constituent parts of both linguistic and rhythmic syntactic processing. Therefore, on one hand, it is possible to say that both domains are implemented in the same neural structures and even the same neural circuits. On the other hand, a closer look at the way linguistic syntactic and rhythmic syntactic processing are implemented in those structures and circuits reveals that both domains are processed differently within those overlapping neural structures. As reviewed above, among different parallel circuits, linguistic syntactic processing makes demands on the executive circuit, while rhythmic syntactic processing in music primarily engages the motor circuit. Thus, I hypothesize that linguistic syntax and beat-based encoding of rhythm both rely on the BG and the CBGT circuits, but do so in different ways.

It is worth noting that I do not suggest qualitative difference between both domains in the way that linguistic syntactic processing relies on a particular subpart of the structures and beat-based encoding of rhythm on another. Indeed, though beat-based encoding of rhythm

predominantly calls for the motor circuit, at least one study demonstrates the activation of the caudate together with the ventrolateral prefrontal cortex, indicating activation of the executive circuit rather than the motor circuit (Kung et al., 2013). That is, rhythmic syntactic processing involves both motor and executive circuits with an emphasis on the former. While linguistic syntactic processing relies more on the executive circuit, the putamen is rather involved in motor coordination for speech articulation (Watkins & Jenkinson, 2016) and thus more contributes to phonological syntax. However, as mentioned above, some studies showed the putamen involvement in linguistic syntactic processing (Friederici, Rüschemeyer, et al., 2003; Moreno et al., 2018). In the case of phonological syntactic processing, too, both the executive and the motor loop are involved (Bohland, Bullock, & Guenther, 2010; Guenther, 2016). Therefore, I regard linguistic syntactic processing (including phonological syntax) and rhythmic syntactic processing as different uses of the same neural structures, or rather as varying degrees of expression on the motor to cognitive gradient (*Figure 15.2*).

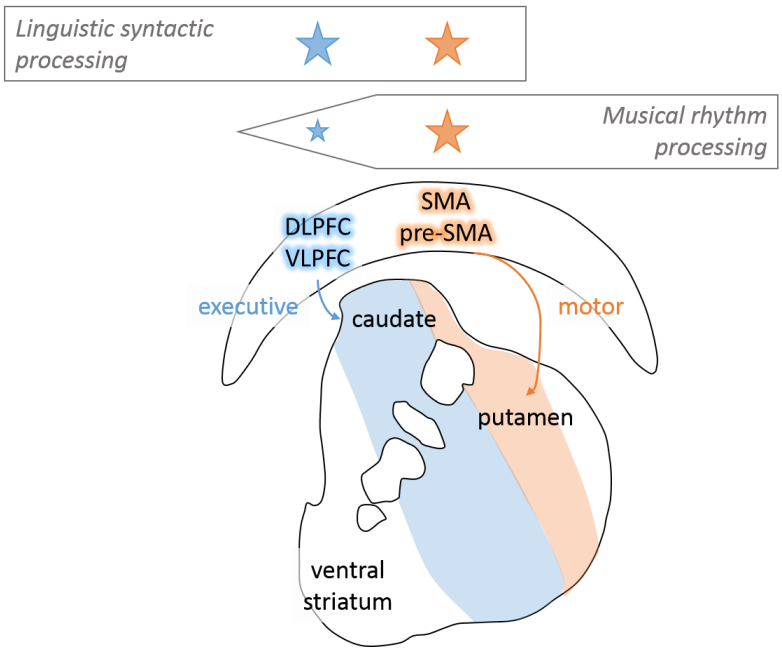


Figure 15.2 Linguistic and rhythmic syntactic processing as varying degrees of expression on the motor to cognitive gradient within the BG and the CBGT circuits. Linguistic syntax processing engages the caudate and the executive circuits (the cortico-striatal projection blue area) to a great extent, while rhythmic syntax rather makes use of the putamen and the motor circuit (the cortico-striatal projection red area). Musical rhythm processing extends to the caudate, but only to a minor degree. DLPFC (dorsoslateral prefrontal cortex); pre-SAM (pre-supplementary motor area); SMA (supplementary motor area); VLPFC (ventrolateral prefrontal cortex).

The investigations of the CBGT circuits provide an opportunity to explore how language and music emerge based on the same mechanisms in terms of "modifiable cortico-striatal synapses" (Dominey & Inui, 2009, p. 1014) and "specialized sub-circuits subserve different domains" (Ullman, 2006, p. 483). Several authors pointed out the functions of the BG and the executive CBGT circuits in learning of grammatical and procedural knowledge (Krishnan, Watkins, & Bishop, 2016; E. K. Miller & Buschman, 2008; Ullman, 2006; Ullman & Pierpont, 2005). In particular, Miller and Buschman (2008) suggested two kinds of learning mechanisms supported by different neural mechanisms. The first one is rapid reinforcement learning by the BG, which leads to acquisition of concrete rules associating specific experiences to rewards. This type of learning is guided by dopaminergic modulatory processing. The second one is the slower learning by the prefrontal cortex (PFC), which facilitates generalized rule acquisition by detecting the regularities and commonalities across many different experiences. Thus, the CBGT circuits, in which those mechanisms form a closed loop, allow for the 'bootstrapping'; i.e. "the process of building increasingly complex representations from simple ones" (E. K. Miller & Buschman, 2008, p. 433).

In this line, regarding impaired syntactic processing ability of the SLI children, the BG role in rule learning was hypothesized (Krishnan et al., 2016; Ullman & Pierpont, 2005). One systematic review article discussing neuroimaging studies about childhood language disorders found out that several studies consistently identified atypical brain structures in the inferior frontal gyrus (especially in the pars triangularis), the caudate nucleus, and the posterior superior temporal gyrus (Mayes, Reilly, & Morgan, 2015). Concerning the SLI, the anomaly in the caudate is also pointed out by Ullman and Pierpont, (2005) as well as Krishnan, Watkins, and Bishop (2016). Some SLI children also show deficits in rhythmic syntactic processing (Corriveau & Goswami, 2009; Cumming, Wilson, Leong, Colling, & Goswami, 2015). This is one additional evidence for the relationship between linguistic and rhythmic syntactic processing in the CBGT circuits.

The CBGT circuits are involved in and are necessary for *learning and performing procedures*, i.e. organizing sequences of actions toward a goal. Especially, the basal ganglia are known to contribute as a 'control center' and an 'associative learner' in both motor and cognitive domains. For example, they play a part in two aspects of motor control: 1) assisting execution of cortically driven predictable and automatic motor patterns; and 2) adapting to unusual circumstances by interrupting and altering the automatically running motor representations (Marsden & Obeso, 1994). Those dual motor functions of the basal ganglia facilitate flexibility and adaptation in motor control. The CBGT circuits are also involved in

and necessary for cognitive control, i.e. control of lower level sensory, memory and/or motor operations to achieve a goal (Graybiel, 1997; Lieberman, 2002, 2016; E. K. Miller, 2000; E. K. Miller & Cohen, 2001; Monchi et al., 2001). Moreover, the basal ganglia contribute to goal-directed (or reward-based) learning processes by acquiring rules associating signal, action, and goal, reinforcing new behavior-guiding rules, and predicting a goal (Graybiel & Mink, 2009; Haber, 2003; Haruno & Kawato, 2006; Lieberman, 2016). As computational goals of language and music are different, those cognitive systems might be canalized differently within the CBGT circuits during the development.

In sum, the following hypotheses can be derived: 1) the putamen-based circuits assist execution of cortically driven predictable and automatic motor and cognitive patterns; 2) the caudate-based circuits get involved in adapting to unusual circumstances where prediction errors happen frequently by interrupting and altering the automatically running motor and cognitive representations. Those hypotheses are in line with general motor and cognitive functions, language syntactic functions as well as rhythmic syntactic functions of the BG and the CBGT circuits. Moreover, those hypotheses imply that, in both language and music, experiments requiring engagement and maintenance of internal periodic prediction activate the putamen-based circuits (Type 1), while deviation from the regularity activates the caudate-based circuits (Type 2).

PART IV Conclusions and future directions

16 Main results of the current thesis

16.1 Comparative biomusicology as a comparative biological information processing framework

The current thesis tackled the question "Why is music the way it is?" within the framework of *comparative biomusicology* by focusing on musical syntax and its relation to syntax in language. *Comparative biomusicology* was introduced as a framework investigating biological foundations of music by integrating 1) comparative approaches, 2) biological approaches and 3) information processing approaches. Comparative approaches include different strategies such as within-domain comparisons (e.g. cross-culture comparison), between-domain comparisons (e.g. language, music, and action), within-species comparisons (e.g. (developmental) disorders), as well as between-species comparisons (e.g. non-human animals and humans). Biological approaches are based on Tinbergen's four questions, i.e., causation / mechanism, ontogeny, evolution / phylogeny, and function. Information processing approaches, which are at the core of cognitive science, contain different levels of analysis known as David Marr's three levels of analysis, i.e., the computational level, the algorithmic level, and the implementational level. Therefore, *comparative biomusicology* as an integrative approach provides a comparative biological information processing framework.

At the very heart of the current thesis lies the question of causation / mechanism, which should be investigated at the algorithmic and implementational level. Within the framework of *comparative biomusicology*, mechanism plays a mediator role for biology and information processing. Identification of computational problems / functions also played a key role as it is a necessary step in order to approach mechanistic explanations of cognitive systems. Computational problems and functions provide top-down constraints to the mechanisms, and in that way they determine what problems the mechanisms have to solve and why. One goal of comparative approaches is to clarify computational problems and mechanisms of cognitive systems such as music and language by applying the four comparative strategies mentioned above. Ontogeny and phylogeny were regarded as changes in mechanism (i.e., algorithms and implementations) and/or function (i.e., computational problems to be solved). The main content of the current thesis did not touch on ontogeny and phylogeny, but I will introduce some implications later in this chapter. In particular, I discussed the following three questions in the thesis: 1) What are the computational problems of music, and musical syntax in particular?; 2) What are neurocognitive mechanisms underlying music syntactic processing?; and 3) How do computational problems and neurocognitive mechanisms of musical syntax relate to those of language?

As starting points, I suggested linking sound and affect as a computational problem for music, and tonal encoding of pitch and beat-based encoding of rhythm as two constituent parts of music to be investigated at the algorithmic level in terms of parsing as well as at the implementational level in terms of neural implementation. To approach a mechanistic explanation in terms of cognitive and neural processes, I proceeded along the following methodological pipeline. First, some analyses were conducted at the computational level to identify computational problems. Second, the computational level analysis was related to processing, especially in terms of parsing. Third, reviews and meta-analyses were conducted to figure out the neural correlates of the processing. Based on the results concerning processing and neural correlates, candidate cognitive and neural processes have been worked out. In this way, the current thesis provided a first step toward resolving the problem of explanatory gaps including a mapping from theoretical concepts of music theory (and linguistic theory) onto hypothetical constructs of psychology and a mapping between mind and brain.

Concerning the relationship between language and music, I introduced a possible way to achieve a non-contrastive comparative approach, namely principled explanations, i.e. regarding cognitive systems as different instances of the same principles. As such, the question "Why is music the way it is?" was investigated in terms of how a set of basic principles is 'put into use' in the way music does. In particular, the current thesis was concerned with syntax as a set of combinatorial principles generating hierarchically structured representations and/or sequences, and neural structures and operations as implementational principles. Thus, instead of emphasizing differences between language and music in details, I largely focused on clarifying the principles underlying language and music that can be abstracted at different levels of comparative inquiry. Then, I discussed to what extent differences between language and music can be explained in terms of different instances of the basic abstract principles revealed so far. Finally, I provided a possible mechanistic explanation stating that language and music processing can be partly explained in terms of the same neurocognitive mechanisms with different expressions on the motor-to-cognitive gradient. In this way, I provided principled explanations to language and music.

16.2 Computational problem of musical syntax: Mapping between hierarchical structure and temporal sequence to link sound and affect

One of the computational problems to be solved by musical syntax, i.e., a core function of syntactic computation, is the mapping between hierarchical structure and temporal sequence to link sound and affect. Musical syntax at least includes two subcomponents, namely tonal-

harmonic syntax (PART II) and rhythmic syntax (PART III). Tonal-harmonic syntax links sound and affect on the basis of tonal hierarchy which represents stability relationships between pitches. Relative importance of pitches is determined in relation to the most stable and important pitch, i.e., the tonic. Rhythmic syntax links sound and affect on the basis of hierarchical metrical and grouping structure. Metrical structure represents relative prominence of beats, while grouping structure represents musical units such as motives, phrases, and sections. Relative salience of beats is determined by interaction between metrical and grouping structure. A beat is regarded as stable if a phenomenal accent matches with relative salience of the beat and as unstable in case of mismatch. In both tonal-harmonic and rhythmic syntax, a transition from relatively stable to unstable pitch creates tension, while a transition from relatively unstable to stable pitch causes relaxation. In addition, ambiguity, i.e., parallel existence of multiple possible representations, also leads to tension.

Concerning the relationship between syntax in language and music at the computational level, this characterization of the computational problem adds to non-contrastive comparative research: One of the computational problems to be solved by syntax in language is likewise the mapping between hierarchical structure and temporal sequence to link sound and meaning. Although the computational problem of language, i.e., linking sound and meaning, differs from that of music, i.e., linking sound and affect, at least one of the computational problems for syntax is the same for language and music, i.e., mapping hierarchical structure and temporal sequence. Of course, syntax in language differs in many ways from tonal-harmonic and rhythmic syntax. However, syntax in language and music can be comparatively investigated in terms of mapping between hierarchical structure and temporal sequence. Then, we can start discussing whether the same formalism (e.g., formal grammar, merge, constructions, or schemas) provides an appropriate characterization of the computational problem for syntax in language and music.

Recursion and cyclicity were suggested to belong to a set of abstract principles according to which combinatorial and labeling operations should perform the mapping between hierarchical structure and temporal sequence. Of course, they are not exhaustive. However, those principles provide important constraints to computational-representational theories and the other levels of investigation. Whatever formalism is applied to characterize the computational problem for syntax in language and music, computational-representational theories of language and music might not be different in those principles. Moreover, those principles especially determine the nature of the representations processed by the algorithms. One example in language are sentences with embedded relative clauses: How does an algorithm

solve the problem of mapping hierarchical structure with multiple embedded clauses and temporal sequence to correctly link sound and meaning? An example in music is syncopation: How does an algorithm solve the problem of mapping hierarchical metrical structure and temporal sequence to meaningfully link sound and affect?

Importantly, this does not mean that there is only one algorithm which operates according to the principles at the computational level because the same computational problem can be solved by different algorithms. Although all levels of investigation relate to each other, it is still central to differentiate the question of *what* the device does and *how*, i.e., the functional question of what task has to be solved at the computational level and mechanistic question at the algorithmic and implementational levels.

16.3 Neurocognitive mechanisms of music syntactic processing revealed by investigating tonal encoding of pitch and beat-based encoding of rhythm

To approximate mechanistic explanations to musical syntax in terms of cognitive and neural processes, the current thesis dealt with tonal encoding of pitch and beat-based encoding of rhythm at the algorithmic and implementational levels. Concerning the algorithmic level, I regarded music syntactic parsing as online execution of mapping between hierarchical structure and temporal sequence to link sound and affect, and investigated it in terms of representations and cognitive processes to identify how music syntactic parsing is performed.

Tonal encoding of pitch contains transformation of a sequence of pitch events into a hierarchical structure representing their relative stability based on a tonal hierarchy. I regarded this aspect of tonal encoding of pitch as tonal-harmonic syntactic processing. This transformation requires sub-processes such as extracting a tonal center, establishing structural relationship, hierarchical structure building as well as structural reanalysis and repair. Thus, experimental studies which attempt to examine tonal-harmonic syntactic processing manipulate the stimuli and tasks in terms of tonal hierarchy and the aforementioned sub-processes. Extracting a tonal center is tested in terms of tonality (in contrast to atonality), establishing structural relationships and hierarchical structure building in terms of grammaticality, regularity, relatedness, and tension-relaxation, and structural repair and reanalysis in terms of modulation.

I carried out an ALE meta-analysis of fMRI studies investigating tonal-harmonic syntactic processing which revealed the left anterior superior temporal gyrus (BA 22) and a right fronto-temporal network including the posterior superior / middle temporal gyrus (BA 22), the inferior frontal gyrus (BA 44 and BA 47), the middle frontal gyrus (BA 9 and 46) in

addition to the right insula. Although this meta-analysis did not show significant activation in the left BA 44, several neuroimaging studies emphasize its importance in establishing structural relationship and hierarchical structure building. An analysis conducted by plotting peak coordinates of those studies revealed that the activation peaks are especially pronounced in the ventral part of BA 44. Therefore, the role of the left BA 44 for tonal-harmonic syntactic processing is still controversial. More fMRI studies examining tonal-harmonic syntactic processing with more variable stimuli and tasks to manipulate different sub-processes are desiderata for future research.

Beat-based encoding of rhythm includes transformation of a sequence of auditory events into a hierarchical structure representing their relative salience of beats. I regarded this aspect of beat-based encoding of rhythm as rhythmic syntactic processing. This transformation requires sub-processes such as beat abstraction (and generation), processing structural relationship, and processing structural ambiguity. Thus, again, experimental studies which attempt to examine rhythmic syntactic processing manipulate the stimuli and tasks in terms of metrical and grouping structures as well as those sub-processes. Beat abstraction is tested in terms of beat-based rhythmic stimuli (in contrast to non-beat rhythm, for example), processing structural relationship in terms of meter and grouping as well as syncopation, and processing structural ambiguity in terms of polyrhythm and drum break.

Another ALE meta-analysis of fMRI studies investigating rhythmic syntactic processing carried out for this thesis revealed significant activations in the supplementary motor area (SMA) (BA 6), the right basal ganglia (putamen), the left cerebellum (culmen), and the left anterior superior temporal gyrus (BA 22). That is, the timing networks including the motor cortico-basal ganglia-thalamocortical (CBGT) circuit and the cerebellum together with the left auditory association area indicate an interaction between auditory and motor systems to yield rhythmic syntactic processing. Although my meta-analysis did not show any significant inferior frontal gyrus activation, some studies demonstrated activations in the bilateral BA 44, 45, 46, and 47. Here, too, more fMRI studies investigating rhythmic syntactic processing in terms of subcomponents introduced in the current thesis are required to reveal the role of the inferior frontal gyrus for rhythmic syntactic processing.

Thus, by investigating tonal-harmonic and rhythmic syntactic processing in terms of their subsystems, it was possible to identify their candidate neural correlates. However, tonal-harmonic syntactic processing, rhythmic syntactic processing, and their sub-processes are cognitive processes which don't have an appropriate granularity to be implemented in the brain as neural processes and, thus, to achieve a mechanistic explanation in terms of cognitive and

neural processes as pursued in the current thesis. Although the specification of sub-processes alongside computational-representational theories is a necessary first step, a purely top-down approach has its limit if the target of research comprises neurocognitive mechanisms. Thus, the current thesis introduced a complementary bottom-up approach taking neural structures and operations as implementational principles. More specifically, this approach was elaborated by focusing on the CBGT circuits.

The CBGT circuits carry out processes such as maintenance and control which can be interpreted as cognitive and neural processes. Candidate neural implementations of the maintenance function comprise delay neurons in the prefrontal cortex as well as the recurrent excitation of the prefrontal pyramidal neurons. The control function seems to be implemented by interaction between cortical and basal ganglia activities. Especially, the basal ganglia control cortical representations by integration, selection, and modulation functions. First, the medium spiny neurons (MSNs) in the striatum serve as integrators by detecting coincidental inputs from cortical areas. Second, by inhibiting or releasing cortical representations, the basal ganglia are able to select a salient representation. The selection function of the basal ganglia is carried out based on the winner-take-all principle. Third, the dopaminergic input from the substantia nigra pars compacta into the striatum weights cortical inputs differently by modulating the firing thresholds of the MNSs.

In the context of music syntactic processing, those cognitive and neural processes were discussed by focusing on rhythmic syntactic processing. For example, I suggested that the motor CBGT circuit involving the SMA and the putamen plays an important role in maintaining periodic predictions because the SMA represents different time intervals and the putamen keeps selecting a particular interval by suppressing the alternatives. In addition, I hypothesized that, especially in case of polyrhythm, multiple representations should be maintained (possibly in BA 47) and that the CBGT circuits might contribute to selecting a currently most salient interpretation by weighting it. The functions of the caudate in rhythmic syntactic processing were hypothesized as generating the beat in concert with external cues by integrating different sources of information and processing prediction errors, possibly with a strong collaboration with the cerebellum.

16.4 Neurocognitive mechanisms for language and music syntactic processing: Going beyond the shared/distinct dichotomy

Ever since the resource sharing framework for syntactic processing in language and music was introduced by Patel (2003), especially since two prominent books (Koelsch, 2012a;

Patel, 2008) have been published on the subject matter, there is a general consensus that there are neural resources shared in language and music syntactic processing. However, it is still not clear what those neural resources are. Generative neurolinguistics and neuromusicology, which apply a top-down strategy by directly mapping concepts of computational-representational theories onto the algorithmic and implementational level research, suggested that language and music share hierarchical processing implemented in Broca's region. However, hierarchical processing (even in its sub-processes such as structure building and establishing structural relationship) again is not at an appropriate level of granularity to be implemented in the brain as a neural process. Neurocognitive psycholinguistics and psychomusicology, on the other hand, suggest executive functions such as working memory and/or cognitive control as candidates for shared aspects of language and music syntactic processing. Executive function in its sub-processes such as maintenance, manipulation, coordination, and switching are cognitive processes which could be implemented in the brain as neural processes. Yet, how such processes solve the computational problem for language and music, i.e., mapping hierarchical structure and temporal sequence to link sound and meaning or affect, was not made clear by those approaches.

Furthermore, the hypotheses about shared neural resources for syntactic processing in language and music were mainly tested in terms of interaction and neural overlap as implications on the basis of the following inferences: 1) Sharing → Interaction; 2) Sharing → Overlap. The current thesis suggests that the second inference (i.e., Sharing → Overlap) is a weak one. First, the comparison of the meta-analysis of tonal-harmonic syntactic processing with that of language comprehension revels little overlap. In particular, the fronto-temporal network of language is more strongly articulated in the left hemisphere, while that of tonal-harmonic syntactic processing can mostly be found in the right hemisphere. Second, within Broca's region in which language and music processing were suggested to overlap, there is an area that is more sensitive to language than other domains. Even BA 44 can be separated further into functionally separable clusters with the 'language cluster' being located in the anterior dorsal BA 44 and the 'action and music cluster' in the posterior ventral BA 44. Third, language syntactic processing and rhythmic syntactic processing both rely on the CBGT circuits, whereas the former is based more heavily on the executive circuit, while the latter rather relies on the motor circuit.

Last but not least, "[i]t is important to keep in mind that neural overlap does not necessarily entail neural sharing" (Peretz, Vuvan, Lagrois, & Armony, 2015, p. 3) because the pattern of activation and/or connectivity could be very different for language and music, even

though there is an overlap. This is a standard logical issue. As “Sharing → Overlap” is the first premise of the argument, it is a logical fallacy to infer sharing from overlap. Of course, as discussed in the current thesis, this type of reverse inference is not uninteresting. However, in the current case, the probability of overlapping activation areas given shared cognitive process is far from optimal and even near to zero. Thus, this reverse inference is very weak.

What are cognitive processes shared between language and music syntactic processing which are in such a granularity that they can be implemented in the brain in terms of neural processes? How do those processes account for the computational problem to be solved by syntax in language and music? If overlap is a weak argument for sharing, what is an alternative approach? The current thesis provided an approach to tackle those three questions by working out a mechanistic explanation for the relationship between language and music syntactic processing, i.e., revealing cognitive and neural processes. The central claim, which is at the same time the main result of the current thesis, is that the same neurocognitive mechanisms are differently expressed on the motor-to-cognitive gradient to solve partially different computational problems of language and music syntactic processing. The details of this approach were worked out by focusing on the relationship between language syntactic processing and rhythmic syntactic processing.

Both linguistic syntactic processing and rhythmic syntactic processing require the CBGT circuits. Concerning language syntactic processing, the executive circuit including the prefrontal cortex and the caudate plays a role in suppressing automatic syntactic representation and selecting an alternative one to process, for example, non-canonical sentences. Regarding rhythmic syntactic processing, the motor circuit including the SMA and the putamen acts as a part in suppressing alternative rhythmic representations and continuously activating the current representation to maintain periodic predictions. Both cases particularly require a selection function, which the basal ganglia execute by inhibiting or releasing cortical representations via the thalamus. Therefore, they use the same neurocognitive mechanisms. However, there is a tendency that linguistic syntactic processing rather makes use of the executive circuit, while rhythmic syntactic processing more strongly relies of the motor circuit. Thus, the same neurocognitive mechanisms show the domain-relevant biases on the motor-to-cognitive gradient.

The current thesis also suggests that the motor-to-cognitive gradient is widely applicable beyond the CBGT circuits. There is a rostro-caudal gradient of memory, control, and goal representation in the frontal cortex with motor part in the most caudal part and cognitive or abstract part in the most rostral part. Even within the BA 44 alone, there is a dorsal-

ventral axis for the maintenance function that extends from the general maintenance function in the most dorsal part via the part associated with language to the most ventral 'action and music' part. In those cases, the main idea is the same: language and music make use of the same neurocognitive mechanism of maintenance, for example, but differ in their degrees of expression on the motor-to-cognitive gradient.

The mechanistic explanation provided by the current thesis is an example of principled explanations to language and music from a bottom-up perspective in terms of neural structures and operations as implementational principles. This kind of principled explanations is necessary if we take comparative approaches, cognitive science, and biological framework seriously. Until now, little has been done to investigate language and music in such a principled way. Thus, the current thesis made a very important step toward a new research strategy investigating the biological foundations of cognitive systems from comparative perspectives.

16.5 Answers to further relevant questions

16.5.1 Did the thesis solve the problem of explanatory gaps?

The mechanistic explanation to language and music syntactic processing worked out in the current thesis provides a partial solution to the problem of explanatory gaps comprising two separate mapping problems: a) How is a mapping from theoretical concepts of music theory (and linguistic theory) onto hypothetical constructs of psychology established to get psychologically relevant and testable concepts?; b) How can we deal with the explanatory gap between mind and brain in comparative biomusicology? Concerning the mapping problem a) linguistic and musicological theories deal with computational problems to be solved as well as with representations to be processed. By explaining how cognitive processes such as maintenance and control operate in transforming temporal sequence to hierarchical structure (and vice versa) to link sound and affect or meaning, the relationship between theoretical and psychological concepts can be established. Moreover, as those cognitive processes can be regarded as neural processes implemented in the brain, the mapping problem b) was also tackled.

For example, the maintenance function implemented in the pre-frontal cortex is required to process the relationship between non-adjacent elements and is thus necessary to implement hierarchical processing of language and music. The control function implemented in the CBGT circuits is essential to select the currently most salient representation among different alternatives. This is of particular importance for resolving tension created by conflicting representations in music and interpreting sentential meaning (e.g., who did what to

whom) in language. Moreover, the cortico-striatal synapses are modifiable and the modulation function implemented by dopaminergic weighting from the substantia nigra pars compacta contributes to reinforcement learning. Together with the prefrontal learning mechanism, the modulation function could play a role in building representations by learning, which is still a matter of future investigations.

16.5.2 How does the elementary parts list of the current thesis look like in the end?

In the Part I, I started with a preliminary list of the elementary parts which was elaborated in the course of the current thesis by focusing on syntax and syntactic processing of language and music. *Table 16.1* is a list of the elementary parts as a result of the current thesis, which is again still preliminary and open for changes in future research.

Table 16.1 Elementary parts list (final, but open for changes in future research)

<table>
<tr><th>Language</th><th>Music</th><th>Principles</th></tr>
<tr><td colspan="3">Computational level</td></tr>
<tr><td>Linking sound and meaning</td><td>Linking sound and affect
- Stability condition
- Salience condition</td><td rowspan="2">Recursion
Cyclicity</td></tr>
<tr><td colspan="2">Mapping between hierarchical structure and temporal sequence</td></tr>
<tr><td colspan="3">Algorithmic level</td></tr>
<tr><td colspan="3">– Representations –</td></tr>
<tr><td>Lexical items
Syntactic structure
- Constituent structure
- Syntactic categories
- Serial order
Meaning (e.g., thematic role)</td><td>Tonal encoding of pitch
- Tonal center
- Tonal hierarchy
Beat-based encoding of rhythm
- Beat
- Group
- Metrical structure
- Grouping structure
Event hierarchy
Affect</td><td rowspan="2">Autonomous single syntactic representation
vs.
Parallel representations</td></tr>
<tr><td colspan="2">Headed hierarchical structure and temporal sequence</td></tr>
</table>

<table>
<tr><td colspan="3">– Cognitive processes –
(instead of algorithm – see Section 17.3, p. 178 for discussions)</td></tr>
<tr><td>Syntactic processing
- Structure building
- Agreement checking
- Thematic-role assignment
- Structural reanalysis and repair</td><td>Tonal encoding of pitch
- Extracting tonal center
- Establishing structural relationship
- Hierarchical structure building
- Structural reanalysis and repair
Beat-based encoding of rhythm
- Beat abstraction
- Beat generation
- Processing structural relationship
- Processing structural ambiguity</td><td rowspan="2">Syntax-first principle
vs.
Winner-take-all principle</td></tr>
<tr><td colspan="2">Hierarchical processing and executive function
→ Maintenance and control</td></tr>
<tr><td colspan="3">Implementational level</td></tr>
<tr><td colspan="3">– Neural structure –</td></tr>
<tr><td>Left fronto-temporal network</td><td>Right fronto-temporal network</td><td rowspan="3">Neurons
Neural circuits</td></tr>
<tr><td colspan="2">Left BA 44</td></tr>
<tr><td colspan="2">Cortico-basal ganglia-thalamocortical circuits</td></tr>
<tr><td colspan="3">– Neural operations –
(or neural processes)</td></tr>
<tr><td colspan="2">Maintenance and control
- Maintenance
- (Goal) Representation
- Integration
- Selection
- Modulation</td><td>Delay firing
Recurrent excitation
Coincidence detection
Inhibition
Dopaminergic weighting

Winner-take-all principle</td></tr>
</table>

17 Open questions and future directions

17.1 How valid is it to assess syntax apart from 'semantics'?[65]

The simplest answer is that it is not at all valid to assess syntax apart from 'semantics'. With 'semantics', I indicate its broad understanding which is not limited to propositional meaning of language (for a broad use of "semantics", see Seifert et al., 2013). Concerning language, from a theoretical perspective, as the central computational problem of syntax is mapping between hierarchical structure and temporal sequence to link sound and meaning, there is a tight relationship between syntax and semantics. As for the representations, hierarchical structure of language encodes meaning. That is, in terms of computational problems and representations, the relationship between syntax and semantics is very tight. Research on syntactic processing, too, suggests that it is hard to disentangle syntactic and semantic processing. For example, it is broadly assumed that the left anterior negativity (LAN) stands for syntactic processing, while N400 reflects semantic processing. However, if access to lexical information is required during agreement checking, i.e., a sub-process of syntactic processing, N400 is triggered by agreement mismatch (for a review, see Molinaro et al., 2011).

As the issue is less clear for music, I elaborate the discussion on the relationship between syntax and 'semantics' regarding music below. In doing so, first of all, I stick to the idea that one component of musical 'meaning' is affect, i.e., patterning of tension and relaxation, then extend the discussion to further components. From a theoretical perspective, as the central computational problem of syntax is mapping between hierarchical structure and temporal sequence to link sound and affect, there is a tight relationship between syntax and 'semantics'. As for the representations, hierarchical structure of music encodes affect. That is, again, in terms of computational problems and representations, the relationship between syntax and 'semantics' is very tight. The current thesis showed that both tonal-harmonic and rhythmic syntax demonstrate this tight syntax-affect relationship. Concerning tonal-harmonic syntax, moving away from and back to the tonal center creates a dynamics of tension and relaxation. In rhythmic syntax, tension is caused when metrical and structural accents do not match to phenomenal accent (e.g., syncopation) or when there are multiple conflicting structural interpretations (e.g., polyrhythm).

Several empirical studies provided evidence for the intertwined relationship between syntax and affect in music processing. For example, tension and relaxation patterns encoded in hierarchical structure built on the basis of tonal hierarchy match to listeners' tension ratings

[65] Special thanks to Michael A. Arbib for asking me this question.

well (for reviews, see Krumhansl, 2002; Krumhansl & Lerdahl, 2010). Moreover, harmonically unexpected chords elicit brain responses relating to tonal-harmonic syntactic processing and cause a systematic increase in the electrodermal activity, indicating emotional response (Steinbeis, Koelsch, & Sloboda, 2005, 2006). In addition, N5 event-related potential component elicited in response to unexpected chords got reduced significantly when the simultaneously presented word of a sentence is semantically less expected, indicating that N5 elicited by tonal-harmonic syntactic errors has to do with some 'meaning' related process, i.e., affect processing (Steinbeis & Koelsch, 2008). Those studies show that tonal-harmonic syntactic processing and affect processing strongly relate to each other. Concerning tension and relaxation encoded in rhythm, syncopation is associated with pleasurable engagement with music (Keller & Schubert, 2011; Witek, Clarke, Wallentin, Kringelbach, & Vuust, 2014) although its relation to tension remains unclear (Keller & Schubert, 2011). To the best of my knowledge, there is no empirical study examining affect induced by polyrhythm.

What are the candidate neural correlates of processing musical affect? An fMRI study showed that the left pars orbitalis is sensitive to felt tension indicated by listeners and the right amygdala is more active (Lehne et al., 2013). My meta-analysis of tonal-harmonic syntactic processing conducted in the current thesis showed the activation in the right pars orbitalis (BA 47). Similarly, in musicians, polyrhythm processing shows the bilateral pars orbitalis (BA 47), the supramarginal gyrus (BA 40), and the right anterior insula activation (Vuust et al., 2006) and the left pars orbitalis (BA 47) and the right anterior cingulate (BA 32) (Vuust et al., 2011). Thus, one candidate structure playing a role in processing affect caused by music syntactic structure is the pars orbitalis (BA 47) as an intersection of the inferior frontal gyrus (IFG) and the orbitofrontal cortex (OFC).

In the context of the discussion about music and emotion, the OFC together with the ventromedial prefrontal cortex (VMPFC) was suggested to be involved in high-level, evaluative emotional processing including the integration of internal information (e.g., information on internal states as provided from other affect[66] systems) and external / contextual information such as rules and current goals (Koelsch et al., 2015; Salimpoor, Zald, Zatorre, Dagher, & McIntosh, 2015). The OFC, similar to the anterior cingulate cortex (ACC), is a part of the so-called reward CBGT circuits and projects primarily to the rostral striatum including the medial caudate nucleus, the medial and ventral rostral putamen, and the nucleus accumbens (NAcc) (Haber, 2011, 2016). The VMPFC, also a part of the reward CBGT circuits, is strongly

[66] Here, "affect" is used in a very broad sense and not limited to the patterning of tension and relaxation.

connected to the hypothalamus, the amygdala, and the hippocampus as well as projects to the shell of the ventral striatum and the lateral wall of the caudate (Haber, 2011, 2016). To what extent do the reward CBGT circuits contribute to the discussion on the relationship between syntax and affect in music?

One promising framework is a dopaminergic reward prediction framework (for example, see Salimpoor et al., 2015). One experiment combining positron-emission tomography (PET) and fMRI showed that not only experience of pleasure (i.e., rewarding experience) is associated with dopamine release in the striatum, but also anticipation of rewarding experience (Salimpoor, Benovoy, Larcher, Dagher, & Zatorre, 2011). Importantly, this study also showed that experience and anticipation of reward relate to the effect of dopamine in a different part of the striatum: the experience of reward was linked up with activations of the NAcc, while the anticipation of reward was associated with activations of the caudate. This is striking as there are different dopaminergic-striatal pathways: to put it very simplified, the ventral striatum receives dopaminergic inputs largely from the ventral tegmental area together with the dorsal cell groups of the substantia nigra pars compacta (SNc), while the caudate and putamen rather receive the dopaminergic inputs from the other parts of the SNc (for more details, see Haber, 2011, 2014, 2016).

Future research on the relationship between syntax and affect could profit from such a dopaminergic reward prediction framework by explicitly linking psychological concepts such as anticipation, expectation, and prediction to tonal-harmonic syntactic processing and rhythmic syntactic processing and then identifying cognitive and neural processes. It was suggested repeatedly that the types of prediction (e.g., what and when), its modes (e.g., top-down and bottom-up), its levels (e.g., low-level sensory-based prediction and high-level knowledge-based prediction), and its time-scales (e.g., local structure and large-scale structure) should be differentiated (for discussions, see also Koelsch, Vuust, & Friston, 2019; Patel & Morgan, 2017; Rohrmeier & Koelsch, 2012; Seifert et al., 2013; Tillmann, 2012). Do those conceptual differences have consequences for mechanistic explanations and (if yes) how? Research investigating those questions could potentially extend mechanistic explanations for the relationship between language and music beyond syntactic processing.

So far, I discussed ‘meaning’ of music in terms of affect, i.e., patterns of tension and relaxation, encoded on the basis of musical syntax. Importantly, as such, patterns of tension and relaxation do not encode specific emotions, but provide a multivalent quality which can be evaluated in terms of different emotions (Krumhansl & Lerdahl, 2010). The affect encoded in hierarchical structure of music does not refer to an external referent. It is based on internal

reference of musical elements to each other and can thus be called intramusical meaning (Koelsch, 2011b; Koelsch & Moya, 2014). Intramusical meaning is not compositional in the way language meaning is. Rather it is constructive in the sense that global level affect builds on local level affects. Intramusical meaning also does not convey unambiguous information as propositional meaning in language does, but exactly this ambiguity is what makes music to a unique mode of communication (Cross, 2011).

Affect as intramusical meaning is, of course, not the only aspect of musical 'semantics'. For example, Cross (2012) criticized Western culture centered research investigating music as "complex and time-ordered sequences of sonic events varying in pitch, loudness, and timbre that are capable of eliciting emotion" (p. 1). Instead, he emphasized the importance of the socio-intentional, pragmatic, or communicative dimension of musical meaning (Cross, 2011, 2012, 2014). Seifert and colleagues (2013) also emphasize the importance of social cognitive neuroscience research investigating musical meaning as grounded in interactions. How does syntax relate to this kind of 'meaning'? One promising starting point is affective and social (or inter-individual) entrainment (Clayton, 2012; Clayton, Sager, & Will, 2004; Knoblich & Sebanz, 2008; Phillips-Silver, Aktipis, & Bryant, 2010; Phillips-Silver & Keller, 2012; Sebanz, Bekkering, & Knoblich, 2006; Trost et al., 2017).

Musical syntax is often neglected in discussions concerning entrainment. However, as already pointed out, rhythmic syntax accounts for abstraction of the primary beats, which then become the shared beats for musical interaction. Moreover, African drum ensemble performance is a good example of how social interaction encodes affect on the basis of rhythmic syntax. All performers play different rhythmic patterns by referring to 12 beats as a cycle. Thus, the inter-individual entrainment may take place at this level. The existence of multiple conflicting hierarchical structures underlying each rhythmic pattern leads to polyrhythmic tension, but there are moments of relaxation identified as down-beats where most of instruments play together. This kind of co-construction of tension-relaxation patterns can be regarded as musical joint action.

Another related question is which formalism might be appropriate to investigate the relationship between language and music without neglecting 'semantics'. The approach developed in the current thesis takes syntax as a mediator for linking sound and 'meaning' in language and music. An alternative approach is a schema-based approach regarding language and music as ways of interacting with, perceiving, representing, and communicating about the internal and external worlds (e.g., emotions and environments) (Seifert et al., 2013). Seifert and colleagues (2013) introduced anticipatory, perceptual, motor and social schemas and seem

to understand "schema" as a dynamic and adaptive unit roughly corresponding to a domain of interaction such as an object or event for perception and a way to act on an object for action. They also suggested that assemblages of such schemas enable flexible anticipation, perception, action, and social interaction.

However, at least in this article, Seifert and colleagues (2013) don't make explicit how those assemblages are 'tuned' and implemented in language and music. In addition, their approach is semantics-centered and its relation to syntax is not clear. Seifert and colleagues (2013, p. 208) say "we are *prepared to interact* with [the environment] in some reasonably structured fashion". Is this "some reasonably structured fashion" their link to syntax? Or do they intend to integrate syntax, semantics, and pragmatics in Morris' semiotics framework dealing with signification and significance? Seifert (2011) pointed out that the signification process is "rooted in actions and motor behavior of organisms, i.e., for humans in social interaction and gestures" (p. 123). Tension-relaxation patterns encoded by musical syntax are 'embodied' in the posture and gesture of conductors or musicians, and dance movement. Therefore, Jackendoff (1987) hypothesized that "musical structures are placed most directly in correspondence with the level of body representation rather than with conceptual structure" (239). A schema-based approach could profit from research on musical gesture.

A further related framework is construction grammar. In general, constructions are form-meaning pairs, i.e. pairing of a particular syntactic pattern with a particular semantic pattern, and are suggested as integrating syntactic, semantic, and sometimes also phonological information (Arbib & Lee, 2008). Lexical items are an example of constructions. Although construction grammar frameworks seem to contribute to working toward a computational neurolinguistics (for discussions, see Arbib, 2016b), it is still a matter of debates how such construction grammar frameworks can add to research on musical syntax and semantics. What are form-meaning pairs in music? This question strongly relates to the nature of musical lexicon and musical memory. Snyder (2016) proposed that structural regularities in music such as tonal hierarchy and meter as well as established musical forms can be stored in musical long-term memory. He also points out the possibility that affect could influence how and what we remember. This could be one possibility to investigate constructions in music. For example, cadence is an example of form-affect pairs, e.g., tonic-subdominant-dominant-tonic (that is, stable-unstable-less unstable-stable) and tension-relaxation-relaxation (Lerdahl, 2013).

17.2 What might be constituents for a between-species comparative approach to language and music?

There are two central questions relating to between-species comparative approach to language and music. 1) How can animal models contribute to mechanistic explanations of language and music?[67] 2) What makes the human brain differ from that of other animals in being language- or music-ready, i.e. 'equipping' for language and music, and how it got that way? Those two questions are closely related, but should be differentiated because the first question deals with the proximate question and the second with the ultimate question. As suggested by De Waal and Ferrari (2010), "[e]very species, including our own, comes with an enormous set of evolutionarily ancient components of cognition that we need to better understand before we can reasonably focus on what makes the cognition of each species special" (p. 202). Thus, animal models could at least inform research on language and music about what those ancient mechanisms are and how they work. They further ask whether "[…] cognitive specializations [are] due to new capacities or rather to new combinations of old ones" (De Waal & Ferrari, 2010, p. 202). That is, the second, ultimate question can be approached by investigating whether and to what extent language and music can be regarded as new combinations of old mechanisms. Below, I elaborate the discussions concerning question 1) and 2).

Marler (1998, 2000) introduced the distinction between phonological syntax / phonocoding and lexical syntax / lexicoding as a fruitful difference for investigating the relationship between animal signaling, speech, and music. Marler (2000, p. 36) defined "phonological syntax" as "recombinations of sound components (e.g., phonemes) in different sequences (e.g., words), where the components themselves are not meaningful". This differs from lexical syntax, defined as "recombinations of component sequences (e.g. words in the lexicon) into different strings (sentences)" (p. 36) which can be regarded as a narrow-sense syntax. Learned birdsongs, for example, "employ phonocoding to create individual song repertoires numbered in the hundreds" (p. 37). This was an important step as animal signaling, speech, and music were made comparable in terms of phonological syntax / phonocoding. This approach was applied by many scholars to investigate the evolution of speech, language, and music (for reviews, see, for example, Berwick, Beckers, Okanoya, & Bolhuis, 2012; Berwick, Okanoya, Beckers, & Bolhuis, 2011; Fitch, 2010a; Rohrmeier et al., 2015; Rothenberg, Roeske, Voss, Naguib, & Tchernichovski, 2014; ten Cate, 2017; ten Cate & Okanoya, 2012).

[67] Special thanks to Erich Jarvis for convincing me to investigate this type of questions.

This comparative research program led to two different research strategies, which I call "linguistics and musicology of animal signaling" and "evolutionary neuroscience of music and language". The former deals with structure of calls and songs, and is well established in bird song research. Especially, complexity of birdsongs, language, and music have been compared in terms of the Chomsky hierarchy. The questions investigated so far center on "Is animal signaling as complex as language?" which is examined by analyzing signals on the basis of the formal language theory as well as "Are non-human animals able to learn a context-free grammar?" which is tested by using the artificial grammar learning (AGL) paradigm (for reviews, see ten Cate, 2017; Wilson, Marslen-Wilson, & Petkov, 2017). Those questions are contrastive in nature and researchers often come up with the conclusion that birdsongs are not complex enough to be compared with human language. In addition, concerning the AGL experiments, it is not at all clear what cognitive processes work together in what way in different species to solve the task (for discussions, see also ten Cate, 2017). However, this does not mean that animal models cannot add to research on neurocognitive mechanisms of language and music as well as their evolution. This also does not mean that only species with 'similar' phenotypes, e.g., song, speech, and music, can be comparatively studied.

The second strategy, i.e., evolutionary neuroscience, is a bottom-up strategy for comparative investigations of neurocognitive mechanisms, which I suggest to adopt in order to identify constituents for between-species comparative approach. This strategy investigates how a set of neural structures and processes (i.e., implementational principles) brings about species specific phenotypes and makes it possible to compare different species in terms of neural structures and processes. For example, parallel functions of the basal ganglia and the CBGT circuits for vocal learning in songbirds and humans have been pointed out repeatedly (Fitch & Jarvis, 2013; Jarvis, 2004; Petkov & Jarvis, 2012). In addition, those structures underlie motor control in non-human primates and humans (Mendoza & Merchant, 2014). The role of the basal ganglia was even suggested for rule-based coding of self-grooming motor sequences in rats (Aldridge & Berridge, 2003). The dopamine regulation in the basal ganglia also plays important roles for associative learning in rodents, songbirds, non-human primates, and humans (for reviews, see Graybiel & Mink, 2009; Matell & Meck, 2004; Simonyan, Horwitz, & Jarvis, 2012). Thus, in terms of motor control and learning, independently of vocal or non-vocal domains, the basal ganglia and the CBGT circuits can be comparatively investigated in a wide range of species (Petkov & Jarvis, 2012) and the comparative research on those structures opens up rich opportunities for comparative genomics and neurogenetics (Enard, 2011; Scharff & Adam, 2013; Scharff & Petri, 2011; Vernes, 2017).

In this way, evolutionary neuroscience can contribute rich animal models to investigate neurocognitive mechanisms of language and music as well as empirically testable evolutionary hypotheses. On the basis of those animal models, the mechanistic explanations of language and music proposed in the current thesis can be extended. For example, studies investigating genetically modified mice carrying a humanized *Foxp2* gene showed an increase in total dendrite length of the MSNs and the effect of a humanized *Foxp2* on total dendrite length was unique in the CBGT circuits (for reviews, see Enard, 2011; Scharff & Petri, 2011). In addition, mice with a humanized *Foxp2* showed significantly more rapid switching of their behavioral strategy (Schreiweis et al., 2014). It is possible that the increase in total dendrite length of the MSNs led to an enhancement of the integration function and enabled effective sensory-motor integration to flexibly select between alternative representations in humans. Moreover, between-species difference in and effect of a humanized *Foxp2* on dopaminergic modulation in the basal ganglia was pointed out repeatedly (Enard, 2011; Scharff & Petri, 2011; Simonyan et al., 2012).

As discussed in the current thesis, flexible switching between alternative representations yielded by integration, selection, and modulation function in the CBGT circuits is crucial for syntactic processing and rhythmic syntactic processing. Given the rich between-species comparability and some species specificity at the same time, the CBGT circuits and motor control as well as learning are optimal constituents for a between-species comparative approach. In humans, their functions might have extended from motor to cognitive functions. In addition, as the CBGT circuits play important roles in learning, they are also optimal means to investigate ontogeny. Again, mice with a humanized *Foxp2* showed increased synaptic plasticity in the cortico-basal ganglia circuits, as reviewed by Enard (2011) as well as Scharff and Petri (2011), indicating possibly more flexible learning mechanism in humans.

Moreover, neurogenetics also makes a better understanding of developmental speech and language disorders possible (Fisher, 2017). One prominent example is the research on the KE family members carrying a *FOXP2* mutation causing speech and language impairment as well as motor-related impairments. In addition, the members with speech and language impairments were deficient in perception and production of rhythm (but not pitch) (Alcock, Passingham, Watkins, & Vargha-Khadem, 2000), which is striking as the meta-analysis of the current thesis showed that the CBGT circuits' role in rhythmic syntactic processing, but not in tonal-harmonic syntactic processing. Thus, research on the CBGT circuits and the related genes might also reveal how innate learning mechanisms give rise to different cognitive systems language and music by interacting with the environment.

The evolutionary neuroscience strategy together with the mechanistic explanations of language and music yielded by the current thesis fulfill the criteria to be a part of the cognitive biology framework integrating cognition, neural mechanisms, and comparative biology (Fitch, 2014). Within this framework, Fitch (2014) argued against the nature / nurture dichotomy and the associanist[68] / cognitivist opposition. He also suggested individual cells as the basic computational unit, brains as predictive systems, and the ability to infer hierarchical structure as being critical for human cognitive systems. I agree that midbrain dopaminergic reward prediction systems are promising targets of research to bridge the gap between cognition and biology. I am also sympathetic to the idea that individual cells should be regarded as the basic computational unit. I highly agree that the nature / nurture dichotomy and associanist / cognitivist opposition can and should be avoided.

However, I emphasize that neural processes such as integration, selection, and modulation are more central to linking cognition and biology than tree-shaped cell structure as such although I definitely do not deny the importance of cell structures. In addition, the emphasis should be not only on what the single cell does, but also on how the cell orchestrates with other structures in larger networks. For example, just identifying that the MSNs' dendrite length is enhanced in humans does not say anything about cognition. Without identifying which neural process is implemented by the MSNs and what effect the particular structure causes on this neural process, it is not possible to make a link to cognition from a biological perspective. In addition, the formal language theory is one useful approach at the computational level, but the mapping between the computational and implementational level is not one to one. This does not mean that the mapping between the computational and implementational level is impossible. Rather, I suggest, instead of claiming the IFG-as-stack model, it is more fruitful to investigate which neural structures and processes are required to implement a stack-like storage and retrieval mechanism in the IFG. The delay neurons and recurrent excitation, for example, could contribute to the maintenance part.

To promote such an evolutionary neuroscience endeavor, comparative atlases integrating neuroanatomy, neurogenetics, and cognitive functions similar to the Zebra Finch Expression Brain Atlas (ZEBrA) (Oregon Health & Science University, Portland, OR 97239; http://www.zebrafinchatlas.org) and Neurosynth (http://www.neurosynth.org/genes/) are

[68] Fitch (2014) seems to use "associanist" to refer to research investigating associative learning, i.e., learning stimulus-response associations, by classical conditioning paradigm.

useful. Platforms such as the Allen Brain Map (http://portal.brain-map.org/) contributed to develop this line of research bridging between cognition and biology.

Last, but not least, the evolutionary neuroscience strategy together with the mechanistic explanations of language and music can also limit the possible hypothesis space for the evolution of language and music. For example, the above discussion about the CBGT circuits can add to the research on the evolution of beat-based encoding of rhythm[69]. Patel (2006) discussed the ability of beat perception and synchronization (BPS) in terms of innateness, domain-specificity, and human-specificity. To investigate the nature of this distinctive component of music cognition and its relation to language, he introduced "vocal learning and rhythmic synchronization hypothesis" which claims that "having the neural circuitry for complex vocal learning is a necessary prerequisite for the ability to synchronize with an auditory beat" (Patel, 2006, p. 102). In particular, it was suggested that basal ganglia were modified to be more sensitive to auditory-motor coupling because of the need for vocal learning (Patel, 2006, 2008).

Patel and Iversen (2014) extended the vocal learning hypothesis by focusing on the predictive nature of beat induction and introduced the "action simulation for auditory prediction" (ASAP) hypothesis which claims that "simulation of periodic movement in motor planning regions provides a neural signal that helps the auditory system predict the timing of upcoming beats" (p. 1). Moreover, they point out the central role of dorsal auditory pathway connections (i.e., connections through the superior longitudinal fasciculus (SLF), in particular the branch 2 (SLF-II) and the temporo-parietal part of SLF (SLF-tp)) for auditory-motor coupling to predict regular temporal pattern in auditory rhythm (Patel & Iversen, 2014). They also suggest that these connections (especially the SLF-tp connections) are more articulated in humans than in non-human primates due to the emergence of vocal learning in our lineage based on the evolutionary neuroscience research conducted by Rilling and colleagues (2008).

Merchant and Honing (2014) also emphasize the importance of auditory-motor coupling for rhythmic entrainment within their Gradual Audiomotor Evolution (GAE) Hypothesis. Their core assumption is that "the human auditory system has a privileged access to the temporal and sequential mechanisms working inside mCBGT [(motor CBGT); RA] circuit in order to determine the exquisite rhythmic abilities of the Homo sapiens" (Merchant & Honing, 2014, pp. 4–5). They mainly mention that two candidate evolutionary changes in

[69] Here, I prefer to use "beat-based encoding of rhythm" as researchers do not explicitly refer to the computational problem of rhythmic syntax and the rhythmic syntactic processing in terms of its sub-processes which I introduced in the current thesis.

the brain enabled the privileged access of the human auditory system to the motor CBGT circuits: 1) the audio-premotor circuits via extension of the dorsal stream projections connecting the superior temporal auditory areas and the frontal lobe, and 2) audio-basal ganglia circuits, i.e. the projections of the superior temporal areas to the basal ganglia (Merchant & Honing, 2014).

The results of the meta-analysis conducted in the current thesis strongly support the importance of the basal ganglia and the motor CBGT circuit for the evolution of the beat-based encoding of rhythm, but not the fronto-temporal network via the SLF. Thus, the changes that happened in the CBGT circuits and their causes seem to be more relevant for the discussion of the evolution of musical rhythm. Concerning the privileged assess of the human auditory system to the motor CBGT circuit, the role of the left anterior superior temporal lobe, which was also consistently activated during the beat-based encoding of rhythm is of interest. In addition, the cerebellum should be definitely integrated to the discussion on the evolution of musical rhythm as the culmen activation was also demonstrated in the current meta-analysis.

17.3 How does a possible way toward computational neurocognitive modeling in comparative biomusicology research look like?

The current thesis examined the neurocognitive mechanisms of music syntactic processing mainly at the implementational level. Concerning the algorithmic level, I worked out the representations and processes of language and music, but did not suggest any concrete algorithm. However, the mechanistic explanations introduced in the current thesis can contribute to develop the algorithms for the computational neurocognitive modeling. Particularly, I suggest that the model developed in the current thesis is a *conceptual* model, i.e., a 'computationally possible' model which can be extended to an explicit *computational* model, i.e., an algorithm written in programming codes to run simulations, of "how a particular brain system operates at […] the level of dynamic interactions in biologically realistic brain regions and neural networks" (Arbib, 2016a, p. 5). Below, I briefly illustrate how such an extension can be made on the basis of the pre-existing explicit computational models and the conceptual model developed in the current thesis.

An example of a computational neurocognitive model potentially fruitful for comparative research on language and music is the ones suggested by Dominey and colleagues (e.g., Dominey, 2005; Dominey & Inui, 2009; Hinaut & Dominey, 2013). They integrated concepts from construction grammar, i.e., theoretical linguistic approach, and results from behavioral, neuroimaging, and neuropsychological experiments in quantitative neurocognitive

models. Those models carry computations, i.e. transforming strings to meaning representations, on the basis of functional hypotheses how specific cognitive processes required for a particular grammatical construction processing, i.e., thematic role assignment, are implemented in the CBGT circuits. The simulation of neuroimaging and neuropsychological studies conducted on the basis of those models showed that the simulation results are comparable to those of the real experimental studies (Dominey, 2005; Dominey & Inui, 2009; Hinaut & Dominey, 2013; Szalisznyó, Silverstein, Teichmann, Duffau, & Smits, 2017).

The models developed by Dominey and colleagues can be regarded as computational neurocognitive models of learning and manipulating sequences regarding serial, temporal, and abstract structures. Thus, it might be possible to extend them to music processing, especially rhythmic syntactic processing, too. However, there should be some difference in the architecture of language and music models as the thematic role assignment recruits the executive loop, while rhythmic syntactic processing rather claims the motor circuits. For example, rhythmic syntactic processing requires time representations in the SMA instead of different classes of word representations in the PFC. Instead of the thematic roles, the metrical structure should be inferred in rhythmic syntactic processing.

Another model which relies more heavily on neural processes implemented in the CBGT circuits is a cognitive control model performing the Wisconsin Card Sorting Task (WCST) and simulating human behavioral data (Caso & Cooper, 2001). This task requires switching between three behavioral strategies, i.e., sorting cards according to the color, shape, or number of the symbols, depending on the current context, i.e., positive or negative feedback. In the model, each behavioral strategy has a dynamically changing activation level. According to the currently most salient behavioral strategy, the current action gets selected. This competition between conflicting behavioral strategies was suggested to be implemented in the CBGT circuits, and thus the authors interpreted their model as the CBGT circuits' model of cognitive control. This mode's algorithm is based on integration, selection, and modulation which are neural processes carried out in the CBGT circuits.

Similar models could be built to investigate flexible switching between alternative representations in language and music. In speech motor control research, Guenther and colleagues developed computational neurocognitive models called DIVA (Directions Into Velocities of Articulators) (Guenther, 2006) and GODIVA (Gradient Order DIVA) (Bohland et al., 2010) with an explicit link to neural structures and processes (for a review, see also Guenther, 2016). Especially, the integration of both models as done by Civier and colleagues (2013) to simulate speech dysfluency in stuttering is of particular interest as it implements the

motor and executive CBGT circuits for selection, initiation, and sequencing of speech movements. This model carries out selection of the currently most salient syllable, i.e., syllable with the highest activation level, among conflicting alternatives on the basis of inhibitory and excitatory control of the CBGT circuits in which cell level processes are also included. This model could be detailed enough to implement neural processes hypothesized in the current thesis concerning rhythmic syntactic processing and abstract enough to extend to non-linguistic domain.

Although the question of how rhythmic syntactic processing can be realized in an algorithm still remains open, the current mechanistic explanation provides rich opportunity for the future computational neurocognitive modeling research. Arbib (2016a) emphasized the importance of understanding the relationship between models and empirical data in terms of 1) empirical data used to constrain the design of the model, 2) those used to test the model's performance, and 3) the simulation results. Those three aspects are integrated in a neuroinformatics database called BODB (Brain Operation Database) as Brain Operation Principles (BOPs), Summaries of Empirical Data (SEDs), and Summaries of Simulation Results (SSRs). One example of the BOPs is the winner take all principles according to which the selection function implemented in the CBGT circuits operate. Such an integrated database could promote the communication between researchers interested in cognition, biology, and computation.

18 Call for comparative cognitive biology

The current thesis introduced *comparative biomusicology* as a comparative biological information processing framework and examined language and music at different levels of investigations. Those levels involved computation, representation and algorithm in terms of parsing, mechanisms, neural implementations in terms of structures and operations, and neural and behavioral measures as demonstrated in Figure 6.1 of PART I. PART II approached the mechanistic explanation by a purely top-down approach from the computational-representational theory via parsing theory. PART III took both top-down and bottom-up strategies and figured out cognitive and neural processes that facilitate language and music syntactic processing. The thesis identified a set of principles in terms of syntax and neural implementations, and yielded principled explanations regarding language and music as different instances of the same principles. At the mechanistic level, in particular, the same neurocognitive mechanisms are differently expressed on the motor-to-cognitive gradient to solve partially different computational problems of language and music syntactic processing.

The mechanistic explanation provided by the current thesis is an example of principled explanations to language and music from a bottom-up perspective in terms of neural structures and operations as implementational principles.

Once a linking hypothesis between mind and brain is formulated at the mechanistic level in terms of cognitive and neural processes as done in PART III of the thesis by integrating top-down and bottom-up approaches, a framework which puts more emphasis on biology, i.e., the cognitive biology framework (Fitch, 2014), is ready for action. Especially, the implementational principles can be refined and their phylogenic changes can be identified by the evolutionary neuroscience strategy. As emphasized in Section 17.2 (p. 173), a cognitive biology framework should regard not only single neurons, but also neural circuits as basic computational units. It should take neural processes into account to investigate why a particular computation is associated with particular neural structures and how that particular computation is carried out by particular neural structures. Moreover, even though such a bottom-up strategy is helpful, a complementary top-down strategy is necessary to restrict the 'search space' in order to find mechanistic explanations (Asano & Seifert, 2018).

Finally, the emphasis on comparative approaches should still remain essential in the proposed framework for cognitive biology as I think biological foundations of each cognitive system can be best investigated in comparative approaches. Therefore, even though the cognitive biology framework includes comparative approaches as a method – as the title of the article by Fitch (2014) clearly states – I would like to call the framework "comparative cognitive biology" for reference in future research. As the readers might have recognized, comparative approaches keep creating new challenges, but for this very reason, I think, comparative approaches will promote future research on the biological foundations of cognitive systems – the human capacities for music and language.

References

Abdul-Kareem, I. A., Stancak, A., Parkes, L. M., & Sluming, V. (2011). Increased gray matter volume of left pars opercularis in male orchestral musicians correlate positively with years of musical performance. *Journal of Magnetic Resonance Imaging*, *33*(1), 24–32. https://doi.org/10.1002/jmri.22391

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*(06), 529–546. https://doi.org/10.1017/S0140525X13003099

Agawu, K. (2006). Structural analysis or cultural analysis? Competing perspectives on the "standard pattern" of West African rhythm. *Journal of the American Musicology Society*, *59*(1), 1–46.

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and Timing Abilities in Inherited Speech and Language Impairment. *Brain and Language*, *75*(1), 34–46. https://doi.org/10.1006/brln.2000.2323

Aldridge, J. W., & Berridge, K. C. (2003). Basal Ganglia Neural Coding of Natural Action Sequences. In A. M. Graybiel, M. R. Delong, & S. T. Kitai (Eds.), *The Basal Ganglia VI* (pp. 279–287). New York: Springer. https://doi.org/10.1007/978-1-4615-0179-4_28

Alexander, G. E., Delong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, *9*(1), 357–381. https://doi.org/10.1146/annurev.ne.09.030186.002041

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., & Zilles, K. (2010). Broca's Region: Novel Organizational Principles and Multiple Receptor Mapping. *PLoS Biology*, *8*(9), e1000489. https://doi.org/10.1371/journal.pbio.1000489

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Arbib, M. A. (2011). From Mirror Neurons to Complex Imitation in the Evolution of Language and Tool Use*. *Annual Review of Anthropology*, *40*, 257–273. https://doi.org/10.1146/annurev-anthro-081309-145722

Arbib, M. A. (Ed.). (2013). *Language, Music, and the Brain. A Mysterious Relationship*. Cambridge, MA: MIT Press.

Arbib, M. A. (2016a). From Neuron to Cognition: An Opening Perspective. In M. A. Arbib & J. Bonaiuto (Eds.), *From Neuron to Cognition via Cognitive Neuroscience* (pp. 1–72). Cambridge, MA: The MIT Press.

Arbib, M. A. (2016b). Towards a Computational Comparative Neuroprimatology: Framing the language-ready brain. *Physics of Life Reviews*, *16*, 1–54. https://doi.org/10.1016/j.plrev.2015.09.003

Arbib, M. A., & Lee, J. (2008). Describing visual scenes: Towards a neurolinguistics based on construction grammar. *Brain Research*, *1225*, 146–162. https://doi.org/10.1016/j.brainres.2008.04.075

Arom, S. (2000). Prolegomena to a Biomusicology. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 27–29). Cambridge, MA; London: The MIT Press.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Asano, R. (2012). Protolanguage, discrete infinity and interfaces: investigating the evolution of the language and music faculty within the minimalist program. In T. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. Hurford (Eds.), *The Evolution of Language: Proceeding of the 9th International Conference (EVOLANG9)* (pp. 26–33). New Jersey: World Scientific.

Asano, R., & Boeckx, C. (2015). Syntax in language and music: What is the right level of comparison? *Frontiers in Psychology*, *6*, 00942. https://doi.org/10.3389/fpsyg.2015.00942

Asano, R., & Seifert, U. (2018). Commentary: The Evolution of Musicality: What Can Be Learned from Language Evolution Research? *Frontiers in Neuroscience*, *12*, 640. https://doi.org/10.3389/fnins.2018.00640

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*(9), 659–669. https://doi.org/10.1038/nrn2667

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences*, *22*(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Bahlmann, J., Gunter, T. C., & Friederici, A. D. (2006). Hierarchical and Linear Sequence Processing: An Electrophysiological Exploration of Two Different Grammar Types.

Journal of Cognitive Neuroscience, *18*(11), 1829–1842. https://doi.org/10.1162/jocn.2006.18.11.1829

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage*, *42*(2), 525–534. https://doi.org/10.1016/j.neuroimage.2008.04.249

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(04), 723–767. https://doi.org/10.1017/S0140525X97001611

Bateson, P., & Laland, K. N. (2013). On current utility and adaptive significance: a response to Nesse. *Trends in Ecology & Evolution*, *28*(12), 682–683. https://doi.org/10.1016/j.tree.2013.10.009

Bechtel, W., & Shagrir, O. (2015). The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322. https://doi.org/10.1111/tops.12141

Bengtsson, S. L., Ullén, F., Henrik Ehrsson, H., Hashimoto, T., Kito, T., Naito, E., … Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex*, *45*(1), 62–71. https://doi.org/10.1016/j.cortex.2008.07.002

Benoit, C.-E., Dalla Bella, S., Farrugia, N., Obrig, H., Mainka, S., & Kotz, S. A. (2014). Musically Cued Gait-Training Improves Both Perceptual and Motor Timing in Parkinson's Disease. *Frontiers in Human Neuroscience*, *8*, 494. https://doi.org/10.3389/fnhum.2014.00494

Bernstein, L. (1976). *The unanswered questions: Six talks at Harvard* (3rd ed.). Cambridge, MA: Harvard University Press.

Berwick, R. C., Beckers, G. J. L., Okanoya, K., & Bolhuis, J. J. (2012). A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience*, *4*(APR), 1–25. https://doi.org/10.3389/fnevo.2012.00005

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. https://doi.org/10.1016/j.tics.2011.01.002

Bharucha, J. J. (1984). Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. *Journal of Experimental Psychology: General*, *113*(3), 421–425. https://doi.org/10.1037/0096-3445.113.3.421

Bianco, R., Novembre, G., Keller, P. E., Kim, S.-G., Scharf, F., Friederici, A. D., … Sammler, D. (2016). Neural networks for harmonic structure in music perception and action.

NeuroImage, *142*, 454–464. https://doi.org/10.1016/j.neuroimage.2016.08.025

Bickerton, D. (2009). Syntax for non-syntacticians. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax* (pp. 3–14). Cambridge, MA: MIT Press.

Bierwisch, M. (1979). Musik und Sprache: überlegungen zu ihrer Struktur und Funktionsweise. In E. Klemm (Ed.), *Jahrbuch der Musikbibliothek Peters: Aufsätze zur Musik.* (pp. 9–102). Leipzig: Peters.

Bischof, N. (2008). *Psychologie. Ein Grundkurs für Anspruchsvolle* (2nd ed.). Stuttgart: Verlag W. Kohlhammer.

Bocanegra, Y., García, A. M., Pineda, D., Buriticá, O., Villegas, A., Lopera, F., … Ibáñez, A. (2015). Syntax, action verbs, action semantics, and object semantics in Parkinson's disease: Dissociability, progression, and executive influences. *Cortex*, *69*, 237–254. https://doi.org/10.1016/j.cortex.2015.05.022

Boeckx, C. (2010). *Language in Cognition. Uncovering mental structures and the rules behind them*. Malden, MA: Wiley-Blackwell.

Boeckx, C. (2013). Biolinguistics: forays into human cognitive biology. *Journal of Anthropological Sciences*, *91*, 1–28. https://doi.org/10.4436/jass.91009

Boeckx, C. (2014). *Elementary Syntactic Structures*. Cambridge: Cambridge University Press.

Boeckx, C., & Grohmann, K. (2007). The biolinguistics manifesto. *Biolinguistics*, *1*, 1–8.

Boeckx, C., Martinez-Alvarez, A., & Leivada, E. (2014). The functional neuroanatomy of serial order in language. *Journal of Neurolinguistics*, *32*, 1–15. https://doi.org/10.1016/j.jneuroling.2014.07.001

Boeckx, C., & Martins, P. T. (2016). Biolinguistics. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics* (pp. 1–15). Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.20

Boeckx, C., & Theofanopoulou, C. (2018). (Neural) Syntax. In Á. J. Gallego & R. Martin (Eds.), *Language, Syntax, and the Natural Sciences* (pp. 295–315). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316591529.016

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529. https://doi.org/10.1162/jocn.2009.21306

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76. https://doi.org/10.1016/j.bandl.2013.01.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016). The Argument Dependency Model. In

G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 357–369). London: Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00030-4

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: An ERP study. *PLoS ONE*, *9*(5), e97467. https://doi.org/10.1371/journal.pone.0097467

Brase, G. L. (2014). Behavioral science integration: A practical framework of multi-level converging evidence for behavioral science theories. *New Ideas in Psychology*, *33*(1), 8–20. https://doi.org/10.1016/j.newideapsych.2013.11.001

Brennan, J. R., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163–173. https://doi.org/10.1016/j.bandl.2010.04.002

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157–158*, 81–94. https://doi.org/10.1016/j.bandl.2016.04.008

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "Ticktock" of our Internal Clock: Direct Brain Evidence of Subjective Accents in Isochronous Sequences. *Psychological Science*, *14*, 362–366. https://doi.org/10.1111/1467-9280.24441

Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, *111*(2), 323–345. https://doi.org/10.1093/brain/111.2.323

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. https://doi.org/10.1177/0305735611425896

Brown, S., Merker, B., & Wallin, N. L. (2000). An introduction to evolutionary musicology. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 3–24). Cambridge, MA: MIT Press.

Bußmann, H. (2008). *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.

Buzsáki, G., Logothetis, N., & Singer, W. (2013). Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms. *Neuron*, *80*(3), 751–764. https://doi.org/10.1016/j.neuron.2013.10.002

Cameron, D. J., Pickett, K. A., Earhart, G. M., & Grahn, J. A. (2016). The Effect of Dopaminergic Medication on Beat-Based Auditory Timing in Parkinson's Disease. *Frontiers in Neurology*, *7*, 19. https://doi.org/10.3389/fneur.2016.00019

Canseco-Gonzalez, E. (2000). Using the Recording of Event-Related Brain Potentials in the Study of Sentence Processing. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.),

Language and the brain: Representation and processing (pp. 229–266). San Diego: Elsevier Academic Press. https://doi.org/10.1016/B978-012304260-6/50014-1

Caso, A., & Cooper, R. P. (2001). A model of cognitive control in the Wisconsin card sorting test: integrating schema theory and basal ganglia function. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 210–215). Cognitive Science Society.

Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S. J. A., Steinberg, F., & Large, E. W. (2010). Neural Responses to Complex Auditory Rhythms: The Role of Attending. *Frontiers in Psychology*, *1*, 224. https://doi.org/10.3389/fpsyg.2010.00224

Chauvigné, L. A. S., Gitau, K. M., & Brown, S. (2014). The neural basis of audiomotor entrainment: an ALE meta-analysis. *Frontiers in Human Neuroscience*, *8*, 776. https://doi.org/10.3389/fnhum.2014.00776

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Listening to Musical Rhythms Recruits Motor Regions of the Brain. *Cerebral Cortex*, *18*(12), 2844–2854. https://doi.org/10.1093/cercor/bhn042

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008b). Moving on Time: Brain Network for Auditory-Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. *Journal of Cognitive Neuroscience*, *20*(2), 226–239. https://doi.org/10.1162/jocn.2008.20018

Cheung, V. K. M., Meyer, L., Friederici, A. D., & Koelsch, S. (2018). The right inferior frontal gyrus processes nested non-local dependencies in music. *Scientific Reports*, *8*(1), 3822. https://doi.org/10.1038/s41598-018-22144-9

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Chomsky, N. (1957). *Syntactic Structures*. Mouton: The Hague.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1968). *Language and Mind* (3rd ed.). Cambridge: Cambridge University Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origins, and Use*. New York: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*(1), 1–22.

Chomsky, N. (2010). Some simple evo-devo theses: how true might they be for language? In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language: Biolinguistic Perspectives*.

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. *Brain and Language*, *126*(3), 263–278. https://doi.org/10.1016/j.bandl.2013.05.016

Clayton, M. (2012). What is Entrainment? Definition and applications in musical research. *Empirical Musicology Review*, *7*(1–2), 49–56.

Clayton, M., Sager, R., & Will, U. (2004). In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM CounterPoint*, *1*, 1–45.

Clos, M., Amunts, K., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44. *NeuroImage*, *83*, 174–188. https://doi.org/10.1016/j.neuroimage.2013.06.041

Colling, L. J., & Thompson, W. F. (2013). Music, action, affect. In T. Cochrane, B. Fantini, & K. R. Scherer (Eds.), *The Emotional Power of Music: Multidisciplinary perspectives on musical arousal, expression, and social control* (pp. 197–212). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654888.001.0001

Cope, T. E., Grube, M., Singh, B., Burn, D. J., & Griffiths, T. D. (2014). The basal ganglia in perceptual timing: Timing performance in Multiple System Atrophy and Huntington's disease. *Neuropsychologia*, *52*(1), 73–81. https://doi.org/10.1016/j.neuropsychologia.2013.09.039

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, *45*(1), 119–130. https://doi.org/10.1016/j.cortex.2007.09.008

Coull, J. T., Cheng, R.-K., & Meck, W. H. (2011). Neuroanatomical and Neurochemical Substrates of Timing. *Neuropsychopharmacology*, *36*(1), 3–25. https://doi.org/10.1038/npp.2010.113

Cross, I. (2011). The meanings of musical meanings. Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Physics of Life Reviews*, *8*(2), 116–119. https://doi.org/10.1016/j.plrev.2011.05.009

Cross, I. (2012). Cognitive science and the cultural nature of music. *Topics in Cognitive Science*, *4*(4), 668–677. https://doi.org/10.1111/j.1756-8765.2012.01216.x

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music*, *42*(6), 809–819. https://doi.org/10.1177/0305735614543968

Cumming, R., Wilson, A., Leong, V., Colling, L. J., & Goswami, U. (2015). Awareness of

Rhythm Patterns in Speech and Music in Children with Specific Language Impairments. *Frontiers in Human Neuroscience*, *9*, 672. https://doi.org/10.3389/fnhum.2015.00672

De Waal, F. B. M., & Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences*, *14*(5), 201–207. https://doi.org/10.1016/j.tics.2010.03.003

Deutsch, D. (1984). Two issues concerning tonal hierarchies: Comment on Castellano, Bharucha, and Krumhansl. *Journal of Experimental Psychology: General*, *113*(3), 413–416. https://doi.org/10.1037/0096-3445.113.3.413

Dominey, P. F. (2005). From Sensorimotor Sequence to Grammatical Construction: Evidence from Simulation and Neurophysiology. *Adaptive Behavior*, *13*(4), 347–361. https://doi.org/10.1177/105971230501300401

Dominey, P. F., & Inui, T. (2009). Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex*, *45*(8), 1012–1018. https://doi.org/10.1016/j.cortex.2009.03.007

Dowling, W. J. (1984). Assimilation and tonal structure: Comment on Castellano, Bharucha, and Krumhansl. *Journal of Experimental Psychology: General*, *113*(3), 417–420. https://doi.org/10.1037/0096-3445.113.3.417

Durrant, S., Hardoon, D. R., Miranda, E. R., Shawe-Taylor, J., Brechmann, A., & Scheich, H. (2007). Neural correlates of tonality in music. In *Proceedings of NIPS Workshop on Music, Brain and Cognition*. Whistler.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., … Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70–85. https://doi.org/10.1016/j.neuroimage.2016.04.072

Embick, D., & Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: correlational , integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, *30*(4), 357–366. https://doi.org/10.1080/23273798.2014.980750

Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Current Opinion in Neurobiology*, *21*(3), 415–424. https://doi.org/10.1016/j.conb.2011.04.008

Fabb, N., & Halle, M. (2012). Grouping in the stressing of words, in metrical verse, and in music. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, I. Cross, & M. Halle (Eds.), *Language and Music as Cognitive Systems* (pp. 4–21). Oxford: Oxford University Press. https://doi.org/10.1093/acprof

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology : CB*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Memory & Cognition*, *37*(1), 1–9. https://doi.org/10.3758/MC.37.1.1

Fernandez-Duque, D. (2009). Cognitive and neural underpinnings of syntactic complexity. In T. Givón & M. Shibatani (Eds.), *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution* (pp. 433–460). Amsterdam: Benjamins. https://doi.org/10.1075/tsl.85.16cog

Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review*, *22*, 365–380. https://doi.org/10.1515/tlir.2005.22.2-4.365

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language*, *47*(2), 250–272. https://doi.org/10.1016/S0749-596X(02)00004-9

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping*, *24*(2), 79–91. https://doi.org/10.1002/hbm.20070

Fisher, S. E. (2017). Evolution of language: Lessons from the genome. *Psychonomic Bulletin & Review*, *24*(1), 34–40. https://doi.org/10.3758/s13423-016-1112-8

Fitch, W. T. (2005). The Evolution of Music in Comparative Perspective. *Annals of the New*

York Academy of Sciences, *1060*(1), 29–49. https://doi.org/10.1196/annals.1360.004

Fitch, W. T. (2006a). On the Biology and Evolution of Music. *Music Perception*, *24*(1), 85–88. https://doi.org/10.1525/mp.2006.24.1.85

Fitch, W. T. (2006b). The biology and evolution of music: a comparative perspective. *Cognition*, *100*(1), 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Fitch, W. T. (2009). Prolegomena to a Future Science of Biolinguistics. *Biolinguistics*, *3*(4), 283–320.

Fitch, W. T. (2010a). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fitch, W. T. (2010b). Three meanings of "recursion": key distinctions for biolinguistics. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language* (pp. 73–90). Cambridge: Cambridge University Press.

Fitch, W. T. (2011). The Evolution of Syntax: An Exaptationist Perspective. *Frontiers in Evolutionary Neuroscience*, *3*, 9. https://doi.org/10.3389/fnevo.2011.00009

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, *7*, 68. https://doi.org/10.3389/fnsys.2013.00068

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364. https://doi.org/10.1016/j.plrev.2014.04.005

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*, 20140091. https://doi.org/10.1098/rstb.2014.0091

Fitch, W. T. (2016). Dance, Music, Meter and Groove: A Forgotten Partnership. *Frontiers in Human Neuroscience*, *10*, 64. https://doi.org/10.3389/fnhum.2016.00064

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, *24*(1), 3–33. https://doi.org/10.3758/s13423-017-1236-5

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: an overview. *Phil. Trans. R. Soc. B*, *367*, 1933–1955. https://doi.org/10.1098/rstb.2012.0103

Fitch, W. T., Friederici, A. D., & Hagoort, P. (2012). Pattern perception and computational complexity: introduction to the special issue. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1598), 1925–1932. https://doi.org/10.1098/rstb.2012.0099

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377–380.

https://doi.org/10.1126/science.1089401

Fitch, W. T., & Jarvis, E. D. (2013). Birdsong and other animal models for human speech, song, and vocal learning. *Language, Music, and the Brain*, (1871), 499–541.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104. https://doi.org/10.1111/nyas.12406

Fitch, W. T., & Rosenfeld, A. (2007). Perception and production of syncopated rhythms. *Music Perception*, *25*(1), 43–58. https://doi.org/10.1525/MP.2007.25.1.43

Fitch, W. T., & Zuberbühler, K. (2013). Primate precursors to human language: beyond discontinuity. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (pp. 26–48). Oxford: Oxford University Press. https://doi.org/10.1093/acprof

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music*, *42*(2), 190–209. https://doi.org/10.1177/0305735612463949

Foster, N. E. V., & Zatorre, R. J. (2010). A Role for the Intraparietal Sulcus in Transforming Musical Pitch Information. *Cerebral Cortex*, *20*(6), 1350–1359. https://doi.org/10.1093/cercor/bhp199

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Friederici, A. D. (2015). The Neuroanatomical Pathway Model of Language. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 349–356). Amsterdam: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00029-8

Friederici, A. D. (2016). The Neuroanatomical Pathway Model of Language. In G. Hickok &

S. L. Small (Eds.), *Neurobiology of Language* (pp. 349–356). London: Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00029-8

Friederici, A. D. (2017). *Language in Our Brain: The Origins of a Uniquely Human Capacity*. Cambridge, MA: The MIT Press.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, *103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254. https://doi.org/10.1016/j.conb.2012.10.002

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *NeuroImage*, *20*(SUPPL. 1), S8–S17. https://doi.org/10.1016/j.neuroimage.2003.09.003

Friederici, A. D., Kotz, S. A., Werheid, K., Hein, G., & von Cramon, D. Y. (2003). Syntactic comprehension in Parkinson's disease: investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology*, *17*(1), 133–142. https://doi.org/10.1037/0894-4105.17.1.133

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language*, *74*(2), 289–300. https://doi.org/10.1006/brln.2000.2313

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex*, *13*(2), 170–177. https://doi.org/10.1093/cercor/13.2.170

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain Research*, *1146*(1), 50–58. https://doi.org/10.1016/j.brainres.2006.08.038

Frisch, S., Kotz, S. A., von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: the role of the basal ganglia. *Clinical Neurophysiology*, *114*(2), 336–340. https://doi.org/10.1016/S1388-2457(02)00366-8

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing.

Neuroreport, *12*(15), 3391–3394. https://doi.org/10.1097/00001756-200110290-00048

Fritz, J., Poeppel, D., Trainor, L. J., Schlaug, G., Patel, A. D., Peretz, I., … Parsons, L. M. (2013). The Neurobiology of Language, Speech, and Music. *Language, Music, and the Brain*, *10*, 417–459. https://doi.org/10.7551/mitpress/9780262018104.003.0017

Fujioka, T., Fidali, B. C., & Ross, B. (2014). Neural correlates of intentional switching from ternary to binary meter in a musical hemiola pattern. *Frontiers in Psychology*, *5*, 1257. https://doi.org/10.3389/fpsyg.2014.01257

Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. *Journal of Neuroscience*, *35*(45), 15187–15198. https://doi.org/10.1523/JNEUROSCI.2397-15.2015

Fujisawa, T. X., & Cook, N. D. (2011). The perception of harmonic triads: an fMRI study. *Brain Imaging and Behavior*, *5*(2), 109–125. https://doi.org/10.1007/s11682-011-9116-5

Fujita, K. (2016). On the parallel evolution of syntax and lexicon: A Merge-only view. *Journal of Neurolinguistics*, 1–15. https://doi.org/10.1016/j.jneuroling.2016.05.001

Fuster, J. M. (2008a). Overview of Prefrontal Functions : The Temporal Organization of Action. *The Prefrontal Cortex*, 333–385.

Fuster, J. M. (2008b). *The prefrontal cortex*. Amsterdam: Elsevier Academic Press.

García, A. M., Bocanegra, Y., Herrera, E., Pino, M., Muñoz, E., Sedeño, L., & Ibáñez, A. (2017). Action-semantic and syntactic deficits in subjects at risk for Huntington's disease. *Journal of Neuropsychology*. https://doi.org/10.1111/jnp.12120

García, A. M., Sedeño, L., Trujillo, N., Bocanegra, Y., Gomez, D., Pineda, D., … Ibáñez, A. (2017). Language Deficits as a Preclinical Window into Parkinson's Disease: Evidence from Asymptomatic Parkin and Dardarin Mutation Carriers. *Journal of the International Neuropsychological Society*, *23*(02), 150–158. https://doi.org/10.1017/S1355617716000710

Garza Villarreal, E. A., Brattico, E., Leino, S., Østergaard, L., & Vuust, P. (2011). Distinct neural responses to chord violations: A multiple source analysis study. *Brain Research*, *1389*, 103–114. https://doi.org/10.1016/j.brainres.2011.02.089

Gaser, C., & Schlaug, G. (2001). Brain structures differ between musicians and non-musician. *NeuroImage*, *13*(6), 1168. https://doi.org/10.1016/S1053-8119(01)92488-7

Geiser, E., & Kaelin-Lang, A. (2011). The function of dopaminergic neural signal transmission in auditory pulse perception: Evidence from dopaminergic treatment in Parkinson's patients. *Behavioural Brain Research*, *225*(1), 270–275. https://doi.org/10.1016/j.bbr.2011.07.019

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of lingustic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge, MA: The MIT Press.

Gierhan, S. M. E. (2013). Connections for auditory language in the human brain. *Brain and Language*, *127*(2), 205–221. https://doi.org/10.1016/j.bandl.2012.11.002

Gobet, F., Lane, P. C. R., Crocker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243. https://doi.org/10.1016/S1364-6613(00)01662-4

Gordon, R. L., Jacobs, M. S., Schuele, C. M., & McAuley, J. D. (2015). Perspectives on the rhythm-grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences*, *1337*(1), 16–25. https://doi.org/10.1111/nyas.12683

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635–644. https://doi.org/10.1111/desc.12230

Grahn, J. A. (2009). The Role of the Basal Ganglia in Beat Perception. *Annals of the New York Academy of Sciences*, *1169*(1), 35–45. https://doi.org/10.1111/j.1749-6632.2009.04553.x

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics in Cognitive Science*, *4*(4), 585–606. https://doi.org/10.1111/j.1756-8765.2012.01213.x

Grahn, J. A., & Brett, M. (2007). Rhythm and Beat Perception in Motor Areas of the Brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, *45*(1), 54–61. https://doi.org/10.1016/j.cortex.2008.01.005

Grahn, J. A., & Rowe, J. B. (2009). Feeling the Beat: Premotor and Striatal Interactions in Musicians and Nonmusicians during Beat Perception. *The Journal of Neuroscience*, *29*(23), 7540–7548. https://doi.org/10.1523/JNEUROSCI.2018-08.2009

Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cerebral Cortex*, *23*(4), 913–921. https://doi.org/10.1093/cercor/bhs083

Graybiel, A. M. (1997). The Basal Ganglia and Cognitive Pattern Generators. *Schizophrenia Bulletin*, *23*(3), 459–469. https://doi.org/10.1093/schbul/23.3.459

Graybiel, A. M. (2000). The basal ganglia. *Current Biology*, *10*(14), R509–R511. https://doi.org/10.1016/S0960-9822(00)00593-5

Graybiel, A. M., & Mink, J. (2009). The Basal Ganglia and Cognition. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences IV* (pp. 565–585). Cambridge, MA: The MIT Press.

Grossman, M. (1999). Sentence Processing in Parkinson's Disease. *Brain and Cognition*, *40*(2), 387–413. https://doi.org/10.1006/brcg.1999.1087

Grossman, M., Cooke, A., DeVita, C., Lee, C., Alsop, D., Detre, J., … Hurtig, H. I. (2003). Grammatical and resource components of sentence processing in Parkinson's disease: An fMRI study. *Neurology*, *60*(5), 775–781. https://doi.org/10.1212/01.WNL.0000044398.73241.13

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Guenther, F. H. (2016). *Neural control of speech*. Cambridge, MA: The MIT Press.

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy*, *26*(4), 317–330. https://doi.org/10.1016/j.jchemneu.2003.10.003

Haber, S. N. (2011). Neuroanatomy of Reward: A View from the Ventral Striatum. In J. A. Gottfried (Ed.), *Neurobiology of Sensation and Reward* (pp. 235–262). Boca Raton: CRC Press/Taylor & Francis.

Haber, S. N. (2014). The place of dopamine in the cortico-basal ganglia circuit. *Neuroscience*, *12*(282), 248–257. https://doi.org/10.1016/j.neuroscience.2014.10.008.The

Haber, S. N. (2016). Corticostriatal circuitry. In *Dialogues in Clinical Neuroscience* (Vol. 18, pp. 7–21).

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. https://doi.org/10.1016/j.tics.2005.07.004

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*(July), 243–270. https://doi.org/10.3389/fpsyg.2013.00416

Hagoort, P. (2016). MUC (Memory, Unification, Control): A Model on the Neurobiology of Language Beyond Single Word Processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 339–347). Amsterdam: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00028-6

Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*, 199–212. https://doi.org/10.1016/S0926-6410(00)00071-9

Han, S., Northoff, G., Vogeley, K., Wexler, B. E., Kitayama, S., & Varnum, M. E. W. (2013). A cultural neuroscience approach to the biosocial nature of the human brain. *Annual Review of Psychology*, *64*, 335–359. https://doi.org/10.1146/annurev-psych-071112-054629

Harley, T. A. (2001). *The Psychology of Language: from data to theory* (2nd ed.). Hove, New York: Psychology Press.

Harrington, D. L., Zimbelman, J. L., Hinton, S. C., & Rao, S. M. (2010). Neural Modulation of Temporal Encoding, Maintenance, and Decision Processes. *Cerebral Cortex*, *20*(6), 1274–1285. https://doi.org/10.1093/cercor/bhp194

Haruno, M., & Kawato, M. (2006). Different Neural Correlates of Reward Expectation and Reward Expectation Error in the Putamen and Caudate Nucleus During Stimulus-Action-Reward Association Learning. *Journal of Neurophysiology*, *95*(2), 948–959. https://doi.org/10.1152/jn.00382.2005

Hauser, M. D. (2009). The possibility of impossible cultures. *Nature*, *460*(7252), 190–196. https://doi.org/10.1038/460190a

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nature Neuroscience*, *6*(7), 663–668. https://doi.org/10.1038/nn1080

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., … Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology*, *5*, 401. https://doi.org/10.3389/fpsyg.2014.00401

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243–259.

Heinz, J., & Idsardi, W. (2013). What Complexity Differences Reveal About Domains in Language. *Topics in Cognitive Science*, *5*(1), 111–131. https://doi.org/10.1111/tops.12000

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145. https://doi.org/10.1038/nrn3158

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99.

https://doi.org/10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*, 393–402. https://doi.org/10.1038/nrn2113

Hickok, G., & Poeppel, D. (2016). Neural Basis of Speech Perception. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 299–310). London: Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00025-0

Hinaut, X., & Dominey, P. F. (2013). Real-Time Parallel Processing of Grammatical Structure in the Fronto-Striatal System: A Recurrent Network Simulation Study Using Reservoir Computing. *PLoS ONE*, *8*(2), e52946. https://doi.org/10.1371/journal.pone.0052946

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Frontiers in Psychology*, *2*, 112. https://doi.org/10.3389/fpsyg.2011.00112

Hochstadt, J., Nakano, H., Lieberman, P., & Friedman, J. (2006). The roles of sequencing and verbal working memory in sentence comprehension deficits in Parkinson's disease. *Brain and Language*, *97*(3), 243–257. https://doi.org/10.1016/j.bandl.2005.10.011

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*, 85–91. https://doi.org/10.1111/j.1749-6632.2011.06402.x

Honing, H. (Ed.). (2018). *The origins of musicality*. Cambridge, MA: The MIT Press.

Honing, H., Bouwer, F. L., & Haden, G. H. (2014). Perceiving temporal regularity in music: The role of auditory event--related potentials (ERPs) in probing beat perception. In H. Merchant & V. de Lafuente (Eds.), *Neurobiology of interval timing* (pp. 1–39). New York: Springer. https://doi.org/10.1007/978-1-4939-1782-2_16

Honing, H., & Ploeger, A. (2012). Cognition and the evolution of music: pitfalls and prospects. *Topics in Cognitive Science*, *4*(4), 513–524. https://doi.org/10.1111/j.1756-8765.2012.01210.x

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Phil. Trans. R. Soc. B*, *370*, 20140088. https://doi.org/10.1098/rstb.2014.0088

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *Journal of Cognitive Neuroscience*, *18*(4), 665–679. https://doi.org/10.1162/jocn.2006.18.4.665

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: The MIT Press.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, *60*, 653–670. https://doi.org/10.1146/annurev.psych.60.110707.163604

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1991). Musical Parsing and Musical Affect. *Music Perception: An Interdisciplinary Journal*, *9*(2), 199–229.

Jackendoff, R. (2000). Fodorian modularity and representational modularity. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain: representation and processing* (pp. 3–30). San Diego: Elsevier Academic Press.

Jackendoff, R. (2002). *Foundations of language. Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, *1146*, 2–22. https://doi.org/10.1016/j.brainres.2006.08.111

Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception*, *26*(3), 195–204. https://doi.org/10.1525/mp.2009.26.3.195

Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, *87*(3), 586–624. https://doi.org/10.1353/lan.2011.0063

Jackendoff, R. (2015). In Defense of Theory. *Cognitive Science*, 1–28. https://doi.org/10.1111/cogs.12324

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*, 33–72. https://doi.org/10.1016/j.cognition.2005.11.005

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*. https://doi.org/10.1016/j.cognition.2005.04.006

Jacobs, A. M., & Hofmann, M. J. (2013). Neurokognitive Modellierung. In E. Schröger & S. Koelsch (Eds.), *Enzyklopädie der Psychologie: Affektive und Kognitive Neurowissenschaft: C/II/9* (pp. 431–444). Göttingen: Hogrefe Verlag.

Jäger, G., & Rogers, J. (2012). Formal language theory: refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1598), 1956–1970. https://doi.org/10.1098/rstb.2012.0077

Janata, P., & Parsons, L. M. (2013). Neural Mechanisms of Music, Singing, and Dancing. In M. A. Arbib (Ed.), *Language, Music, and the Brain* (pp. 307–328). Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262018104.003.0013

Jantzen, K. J., Oullier, O., Marshall, M., Steinberg, F. L., & Kelso, J. A. S. (2007). A parametric

fMRI investigation of context effects in sensorimotor timing and coordination. *Neuropsychologia*, *45*(4), 673–684. https://doi.org/10.1016/j.neuropsychologia.2006.07.020

Jantzen, K. J., Steinberg, F. L., & Kelso, J. A. S. (2002). Practice-dependent modulation of neural activity during human sensorimotor coordination: a functional Magnetic Resonance Imaging study. *Neuroscience Letters*, *332*(3), 205–209. https://doi.org/10.1016/S0304-3940(02)00956-4

Jantzen, K. J., Steinberg, F. L., & Kelso, J. A. S. (2005). Functional MRI reveals the existence of modality and coordination-dependent timing networks. *NeuroImage*, *25*(4), 1031–1042. https://doi.org/10.1016/j.neuroimage.2004.12.029

Jarvis, E. D. (2004). Learned Birdsong and the Neurobiology of Human Language. *Annals of the New York Academy of Sciences*, *1016*(1), 749–777. https://doi.org/10.1196/annals.1298.038

Jeon, H. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, *8*(November), 1–8. https://doi.org/10.3389/fnsys.2014.00223

Jones, C. R. G., & Jahanshahi, M. (2014). Motor and Perceptual Timing in Parkinson's Disease. In H. Merchant & V. de Lafuente (Eds.), *Neurobiology of interval timing* (pp. 265–290). New York: Springer. https://doi.org/10.1007/978-1-4939-1782-2_14

Jones, Mari Riess. (2016). Musical Time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 125–141). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.13

Kaan, E. (2007). Event-Related Potentials and Language Processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571–591. https://doi.org/10.1111/j.1749-818X.2007.00037.x

Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, *6*(8), 350–356. https://doi.org/10.1016/S1364-6613(02)01947-2

Kaplan, R. M. (2003). Syntax. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (pp. 70–90). Oxford, New York: Oxford University Press.

Katz, J., & Pesetsky, D. (2011). The Identity Thesis for Language and Music. Retrieved from http://ling.auf.net/lingbuzz/000959

Keller, P. E., & Schubert, E. (2011). Cognitive and affective judgements of syncopated musical themes. *Advances in Cognitive Psychology*, *7*, 142–156. https://doi.org/10.2478/v10053-008-0094-0

Kljajevic, V. (2010). Is syntactic working memory language specific? *Psihologija*, *43*(1), 85–101. https://doi.org/10.2298/PSI1001085K

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2021–2031. https://doi.org/10.1098/rstb.2008.0006

Knott, A. (2014). Syntactic Structures as Descriptions of Sensorimotor Processes. *Biolinguistics*, *8*, 1–52.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*(6), 963–974. https://doi.org/10.1016/j.neuron.2006.05.017

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179–190. https://doi.org/10.1111/j.1469-8986.2008.00752.x

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, *14*(3), 131–137. https://doi.org/10.1016/j.tics.2010.01.002

Koelsch, S. (2011a). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, *2*, 110. https://doi.org/10.3389/fpsyg.2011.00110

Koelsch, S. (2011b). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. https://doi.org/10.1016/j.plrev.2011.04.004

Koelsch, S. (2012a). *Brain and Music*. Chichester, West Sussex; Hoboken, NJ: Wiley-Blackwell.

Koelsch, S. (2012b). Response to target article 'Language, music, and the brain: a resource-sharing framework.' In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 224–234). Oxford: Oxford University Press. https://doi.org/10.1093/acprof

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage*, *25*(4), 1068–1076. https://doi.org/10.1016/j.neuroimage.2004.12.050

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541. https://doi.org/10.1162/089892900562183

Koelsch, S., Gunter, T. C., v. Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach Speaks: A Cortical "Language-Network" Serves the Processing of Music. *NeuroImage*, *17*(2), 956–966. https://doi.org/10.1006/nimg.2002.1154

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax

processing in language and in music: an ERP Study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577. https://doi.org/10.1162/089892905774597290

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., & Gebauer, G. (2015). The quartet theory of human emotions: An integrative and neurofunctional model. *Physics of Life Reviews*, *13*, 1–27. https://doi.org/10.1016/j.plrev.2015.03.001

Koelsch, S., & Moya, D. G. (2014). Musical semantics. In W. F. Thompson, A. Lamont, R. Parncutt, & F. A. Russo (Eds.), *Music in the Social and Behavioral Sciences: An Encyclopedia* (pp. 793–795). Los Angeles: Sage.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–15448. https://doi.org/10.1073/pnas.1300272110

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584. https://doi.org/10.1016/j.tics.2005.10.001

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Konoike, N., Mikami, A., & Miyachi, S. (2012). The influence of tempo upon the rhythmic motor control in macaque monkeys. *Neuroscience Research*, *74*(1), 64–67. https://doi.org/10.1016/j.neures.2012.06.002

Kotz, S. A., Brown, R. M., & Schwartze, M. (2016). Cortico-striatal circuits and the timing of action and perception. *Current Opinion in Behavioral Sciences*, *8*(August), 42–45. https://doi.org/10.1016/j.cobeha.2016.01.010

Kotz, S. A., Frisch, S., von Cramon, D. Y., & Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society : JINS*, *9*(7), 1053–1060. https://doi.org/10.1017/S1355617703970093

Kotz, S. A., & Gunter, T. C. (2015). Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Annals of the New York Academy of Sciences*, *1337*(1), 62–68. https://doi.org/10.1111/nyas.12657

Kotz, S. A., & Schmidt-Kassow, M. (2015). Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*, *68*, 48–60. https://doi.org/10.1016/j.cortex.2015.02.021

Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex*, *45*(8), 982–990. https://doi.org/10.1016/j.cortex.2009.02.010

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Krebs, H. (1999). *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann*. New York: Oxford University Press.

Krishnan, S., Watkins, K. E., & Bishop, D. V. M. (2016). Neurobiological Basis of Language Learning Difficulties. *Trends in Cognitive Sciences*, *20*(9), 701–714. https://doi.org/10.1016/j.tics.2016.06.012

Krumhansl, C. L. (2002). Music: A Link Between Cognition and Emotion. *Current Directions in Psychological Science*, *11*(2), 45–50. https://doi.org/10.1111/1467-8721.00165

Krumhansl, C. L., & Lerdahl, F. (2010). Musical Tension. In F. Bacci & D. Melcher (Eds.), *Art and the Senses* (pp. 297–314). Oxford, New York: Oxford University Press.

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS ONE*, *10*(11), 1–16. https://doi.org/10.1371/journal.pone.0141069

Kung, S.-J., Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2013). Interacting Cortical and Basal Ganglia Networks Underlying Finding and Tapping to the Musical Beat. *Journal of Cognitive Neuroscience*, *25*(3), 401–420. https://doi.org/10.1162/jocn_a_00325

LaCroix, A. N., Diaz, A. F., & Rogalsky, C. (2015). The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Frontiers in Psychology*, *6*, 1138. https://doi.org/10.3389/fpsyg.2015.01138

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., … Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, *25*(1), 155–164. https://doi.org/10.1002/hbm.20136

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and Effect in Biology Revisited: Is Mayr's Proximate-Ultimate Dichotomy Still Useful? *Science*, *334*(6062), 1512–1516. https://doi.org/10.1126/science.1210879

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., … Mazziotta, J. C. (1997). Automated labeling of the human brain: A preliminary report on

the development and evaluation of a forward-transform method. *Human Brain Mapping*, *5*(4), 238–242. https://doi.org/10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–131. https://doi.org/10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Large, E. W. (2008). Resonating to musical rhythm: theory and experiment. In S. Grondin (Ed.), *Psychology of time* (1st ed., pp. 189–231). Bingley: Emerald. https://doi.org/10.1016/B978-0-08046-977-5.00006-5

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159.

Lasher, M. D. (1981). The cognitive representation of an event involving human motion. *Cognitive Psychology*, *13*(3), 391–406. https://doi.org/10.1016/0010-0285(81)90015-3

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium* (pp. 112–147). New York: Wiley.

Lehéricy, S., Ducros, M., Krainik, A., Francois, C., Van de Moortele, P.-F., Ugurbil, K., & Kim, D.-S. (2004). 3-D diffusion tensor axonal tracking shows distinct SMA and pre-SMA projections to the human striatum. *Cerebral Cortex*, *14*(12), 1302–1309. https://doi.org/10.1093/cercor/bhh091

Lehéricy, S., Ducros, M., Van De Moortele, P.-F., Francois, C., Thivard, L., Poupon, C., … Kim, D.-S. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology*, *55*(4), 522–529. https://doi.org/10.1002/ana.20030

Lehne, M., Rohrmeier, M., & Koelsch, S. (2013). Tension-related activity in the orbitofrontal cortex and amygdala: An fMRI study with music. *Social Cognitive and Affective Neuroscience*, *9*(10), 1515–1523. https://doi.org/10.1093/scan/nst141

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Leow, L.-A., & Grahn, J. A. (2014). Neural Mechanisms of Rhythm Perception: Present Findings and Future Directions. In H. Merchant & V. de Lafuente (Eds.), *Neurobiology of interval timing* (pp. 325–338). New York: Springer New York. https://doi.org/10.1007/978-1-4939-1782-2_17

Lerdahl, F. (1989). Atonal prolongational structure. *Contemporary Music Review*, *4*(1), 65–87. https://doi.org/10.1080/07494468900640211

Lerdahl, F. (2001). *Tonal Pitch Space*. New York: Oxford University Press.

Lerdahl, F. (2009). Genesis and Architecture of the GTTM Project. *Music Perception*. https://doi.org/10.1525/mp.2009.26.3.187

Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In M. A. Arbib (Ed.), *Language, music, and the brain* (pp. 257–272). Cambridge, MA: The MIT Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Levitin, D. (2003). Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*(4), 2142–2152. https://doi.org/10.1016/j.neuroimage.2003.08.016

Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *American Journal of Physical Anthropology*, *119*(S35), 36–62. https://doi.org/10.1002/ajpa.10171

Lieberman, P. (2010). The creative capacity of language, in what manner is it unique, and who had it? In R. K. Larson, V. Deprez, & H. Yamakido (Eds.), *The Evolution of Human Language* (pp. 163–175). Cambridge: Cambridge University Press.

Lieberman, P. (2016). The evolution of language and thought. *Journal of Anthropological Sciences*, *94*, 1–20. https://doi.org/10.4436/jass.94029

Lobina, D. J. (2017). *Recursion. A Computational Investigation into the Representation and Processing ofLanguage*. Oxford: Oxford University Press.

Locke, D. (1982). Principles of Offbeat Timing and Cross-Rhythm in Southern Ev̲e Dance Drumming. *Ethnomusicology*, *26*(2), 217–246.

London, J. (2012a). *Hearing in Time. Psychological Aspects of Musical Meter* (2nd ed.). Oxford: Oxford University Press.

London, J. (2012b). Three Things Linguists Need to Know About Rhythm and Time in Music. *Empirical Musicology Review*, *7*(1–2), 5–11.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The Rhythmic Interpretation of Monophonic Music. *Music Perception*, *1*(4), 424–441. https://doi.org/10.2307/40285271

Longworth, C. E. (2005). The basal ganglia and rule-governed language use: evidence from vascular and degenerative conditions. *Brain*, *128*(3), 584–596. https://doi.org/10.1093/brain/awh387

Lustig, C., Matell, M. S., & Meck, W. H. (2005). Not "just" a coincidence: frontal-striatal interactions in working memory and interval timing. *Memory (Hove, England)*, *13*(3–4), 441–448. https://doi.org/10.1080/09658210344000404

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed

in Broca's area: an MEG study. *Nature Neuroscience*, *4*(5), 540–545. https://doi.org/10.1038/87502

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences*, *106*(20), 8362–8367. https://doi.org/10.1073/pnas.0810928106

Makuuchi, M., Grodzinsky, Y., Amunts, K., Santi, A., & Friederici, A. D. (2013). Processing Noncanonical Sentences in Broca's Region: Reflections of Movement Distance and Type. *Cerebral Cortex*, *23*(3), 694–702. https://doi.org/10.1093/cercor/bhs058

Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *Linguistic Review*, *22*(2–4), 429–445. https://doi.org/10.1515/tlir.2005.22.2-4.429

Marler, P. (1998). Animal communication and human language. In G. Jablonski & L. C. Aiello (Eds.), *The Origin and Diversification of Language. Wattis Symposium Series in Anthropology. Memoirs of the California Academy of Sciences, No. 24* (pp. 1–19). San Francisco: California Academy of Sciences.

Marler, P. (2000). Origins of music and speech: Insights from animals. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 31–48). London: MIT Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: The MIT Press.

Marr, D., & Poggio, T. (1976). From Understanding Computation To Understanding Neural Circuitry. *MIT AI Lab, Memo 357*, 470–488.

Marsden, C. D., & Obeso, J. A. (1994). The functions of the basal ganglia and the paradox of stereotactic surgery in Parkinson's disease. *Brain*, *117*, 877–897.

Martins, M. D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1598), 2055–2064. https://doi.org/10.1098/rstb.2012.0097

Martins, P. T., & Boeckx, C. (2016). What we talk about when we talk about biolinguistics. *Linguistics Vanguard*, *2*(1), 1–15. https://doi.org/10.1515/lingvan-2016-0007

Masataka, N., & Perlovsky, L. (2013). Cognitive interference can be mitigated by consonant music and facilitated by dissonant music. *Scientific Reports*, *3*(1), 2028. https://doi.org/10.1038/srep02028

Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *BioEssays*, *22*(1), 94–103. https://doi.org/10.1002/(SICI)1521-1878(200001)22:1<94::AID-BIES14>3.0.CO;2-E

Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research*, *21*(2), 139–170. https://doi.org/10.1016/j.cogbrainres.2004.06.012

Mayes, A. K., Reilly, S., & Morgan, A. T. (2015). Neural correlates of childhood language disorder: a systematic review. *Developmental Medicine & Child Neurology*, *57*(8), 706–717. https://doi.org/10.1111/dmcn.12714

Mayr, E. (1961). Cause and Effect in Biology. *Science*, *134*, 1501–1506.

Mayville, J. M., Jantzen, K. J., Fuchs, A., Steinberg, F. L., & Kelso, J. A. S. (2002). Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI. *Human Brain Mapping*, *17*(4), 214–229. https://doi.org/10.1002/hbm.10065

McAuley, J. D. (2010). Tempo and Meter. In Mari R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music Perception. Springer handbook of auditory research, Vol. 36*. New York: Springer.

McAuley, J. D., Henry, M. J., & Tkach, J. (2012). Tempo mediates the involvement of motor areas in beat perception. *Annals of the New York Academy of Sciences*, *1252*(1), 77–84. https://doi.org/10.1111/j.1749-6632.2011.06433.x

McDermott, J., & Hauser, M. D. (2005a). Probing the evolutionary origins of music perception. *Annals of the New York Academy of Sciences*, *1060*, 6–16. https://doi.org/10.1196/annals.1360.002

McDermott, J., & Hauser, M. D. (2005b). The Origins of Music: Innateness, Uniqueness, and Evolution. *Music Perception*, *23*(1), 29–60.

McDermott, J., & Hauser, M. D. (2006). Thoughts on an Empirical Approach to the Evolutionary Origins of Music. *Music Perception*, *24*(1), 111–116. https://doi.org/10.1525/mp.2006.24.1.111

Mendoza, G., & Merchant, H. (2014). Motor system evolution and the emergence of high cognitive functions. *Progress in Neurobiology*, *122*, 73–93. https://doi.org/10.1016/j.pneurobio.2014.09.001

Merchant, H., Grahn, J. A., Trainor, L. J., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B*, *370*, 20140093. https://doi.org/10.1098/rstb.2014.0093

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural Basis of the Perception and Estimation of Time. *Annual Review of Neuroscience*, *36*(1), 313–336. https://doi.org/10.1146/annurev-neuro-062012-170349

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in*

Neuroscience, *7*, 274. https://doi.org/10.3389/fnins.2013.00274

Merchant, H., Luciana, M., Hooper, C., Majestic, S., & Tuite, P. (2008). Interval timing and Parkinson's disease: Heterogeneity in temporal performance. *Experimental Brain Research*, *184*(2), 233–248. https://doi.org/10.1007/s00221-007-1097-7

Merker, B. (2002). Music: the missing Humboldt system. *Musicae Scientiae*, *1*(1), 3–21. https://doi.org/10.1177/102986490200600101

Merker, B. (2005). The Conformal Motive in Birdsong, Music, and Language: An Introduction. *Annals of the New York Academy of Sciences*, *1060*(1), 17–28. https://doi.org/10.1196/annals.1360.003

Merker, B. (2014). Groove or swing as distributed rhythmic consonance: introducing the groove matrix. *Frontiers in Human Neuroscience*, *8*, 454. https://doi.org/10.3389/fnhum.2014.00454

Merker, B., Morley, I., & Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music.

Mestres-Missé, A., Turner, R., & Friederici, A. D. (2012). An anterior–posterior gradient of cognitive control within the dorsomedial striatum. *NeuroImage*, *62*(1), 41–47. https://doi.org/10.1016/j.neuroimage.2012.05.021

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Meyer, L., & Friederici, A. D. (2015). Neural Systems Underlying the Processing of Complex Sentences. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 597–606). Amsterdam: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00048-1

Middleton, F. A., & Strick, P. L. (2000). Basal Ganglia Output and Cognition: Evidence from Anatomical, Behavioral, and Clinical Studies. *Brain and Cognition*, *42*(2), 183–200. https://doi.org/10.1006/brcg.1999.1099

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews. Neuroscience*, *1*(1), 59–65. https://doi.org/10.1038/35036228

Miller, E. K., & Buschman, T. J. (2008). Rules through Recursion: How Interactions between the Frontal Cortex and Basal Ganglia May Build Abstract, Complex Rules from Concrete, Simple Ones. In S. A. Bunge & J. D. Wallis (Eds.), *Neuroscience of Rule-Guided Behavior* (pp. 419–440). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195314274.003.0022

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202.

https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *101*(2), 343–352.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9

Miller, N. S., Kwak, Y., Bohnen, N. I., Müller, M. L. T. M., Dayalu, P., & Seidler, R. D. (2013). The pattern of striatal dopaminergic denervation explains sensorimotor synchronization accuracy in Parkinson's disease. *Behavioural Brain Research*, *257*(4), 100–110. https://doi.org/10.1016/j.bbr.2013.09.032

Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., … Bruzzone, M. G. (2008). FMRI/ERP of musical syntax: comparison of melodies and unstructured note sequences. *NeuroReport*, *19*(14), 1381–1385. https://doi.org/10.1097/WNR.0b013e32830c694b

Mobbs, D., Trimmer, P. C., Blumstein, D. T., & Dayan, P. (2018). Foraging for foundations in decision neuroscience: insights from ethology. *Nature Reviews Neuroscience*, *19*(7), 419–427. https://doi.org/10.1038/s41583-018-0010-7

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to 'e'motion. *Social Cognitive and Affective Neuroscience*, *1*(3), 235–241. https://doi.org/10.1093/scan/nsl029

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting Revisited: Distinct Neural Circuits Participating in Different Stages of the Task Identified by Event-Related Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, *21*(19), 7733–7741. https://doi.org/10.1523/JNEUROSCI.21-19-07733.2001

Moreno, A., Limousin, F., Dehaene, S., & Pallier, C. (2018). Brain correlates of constituent structure in sign language comprehension. *NeuroImage*, *167*, 151–161. https://doi.org/10.1016/j.neuroimage.2017.11.040

Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., & Fazio, F. (2001). Syntax and the Brain: Disentangling Grammar by Selective Anomalies. *NeuroImage*, *13*(1), 110–118. https://doi.org/10.1006/nimg.2000.0668

Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and Nonnative Speakers' Processing of a Miniature Version of Japanese as Revealed by ERPs. *Journal of Cognitive*

Neuroscience, *17*(8), 1229–1244. https://doi.org/10.1162/0898929055002463

Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain Activity Associated with Syntactic Incongruencies in Words and Pseudo-Words. *Journal of Cognitive Neuroscience*, *9*(3), 318–329. https://doi.org/10.1162/jocn.1997.9.3.318

Murphy, E. (2015). The brain dynamics of linguistic computation. *Frontiers in Psychology*, *6*, 1515. https://doi.org/10.3389/fpsyg.2015.01515

Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., … Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *NeuroImage*, *117*, 267–283. https://doi.org/10.1016/j.neuroimage.2015.05.020

Myachykov, A., Tomlin, R. S., & Posner, M. I. (2005). Attention and empirical studies of grammar. *The Linguistic Review*, *22*(2–4), 347–364. https://doi.org/10.1515/tlir.2005.22.2-4.347

Narmour, E. (1977). *Beyond Schenkerism : the need for alternatives in music analysis*. Chicago: University of Chicago Press.

Nettl, B. (2000). An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 463–472). London: MIT Press.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research*, *30*(3), 339–364.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281. https://doi.org/10.3758/CABN.5.3.263

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, *4*(10), 906–924. https://doi.org/10.1111/j.1749-818X.2010.00244.x

Oechslin, M. S., Van De Ville, D., Lazeyras, F., Hauert, C. A., & James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex*, *23*(9), 2213–2224. https://doi.org/10.1093/cercor/bhs206

Ojima, S., & Okanoya, K. (2014). The Non-Hierarchical Nature of the Chomsky Hierarchy-Driven Artificial-Grammar Learning. *Biolinguistics*, *8*, 163–180.

Oullier, O., Jantzen, K. J., Steinberg, F. L., & Kelso, J. A. S. (2005). Neural Substrates of Real and Imagined Sensorimotor Coordination. *Cerebral Cortex*, *15*(7), 975–985.

https://doi.org/10.1093/cercor/bhh198

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 2522–2527. https://doi.org/10.1073/pnas.1018711108

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(4), 728–741. https://doi.org/10.1037/0096-1523.16.4.728

Partee, B. H., Ter Meulen, A., & Wall, R. E. (1993). *Mathematical Methods in Linguistics* (Vol. 30). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-2213-6

Parvizi, J. (2009). Corticocentric myopia: old bias in new cognitive sciences. *Trends in Cognitive Sciences*, *13*(8), 354–359. https://doi.org/10.1016/j.tics.2009.04.008

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*(1), 99–104. https://doi.org/10.1525/mp.2006.24.1.99

Patel, A. D. (2008). *Music, language, and the brain*. Oxford, New York: Oxford University Press.

Patel, A. D. (2012). Language, music, and the brain: a resource-sharing framework. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 204–223). Oxford: Oxford University Press. https://doi.org/10.1093/acprof

Patel, A. D. (2013). Sharing and Nonsharing of Brain Resources for Language and Music. In M. A. Arbib (Ed.), *Language, music, and the brain* (pp. 329–355). Cambridge, MA: MIT Press.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*, 57. https://doi.org/10.3389/fnsys.2014.00057

Patel, A. D., & Morgan, E. (2017). Exploring Cognitive Relations Between Prediction in Language and Music. *Cognitive Science*, *41*(S2), 303–320. https://doi.org/10.1111/cogs.12411

Peretz, I. (2013). The Biological Foundations of Music: Insights from Congenital Amusia. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 551–564). London: Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00013-4

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7),

688–691. https://doi.org/10.1038/nn1083

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140090. https://doi.org/10.1098/rstb.2014.0090

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review*, *20*(2), 310–317. https://doi.org/10.3758/s13423-012-0344-5

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*, 12. https://doi.org/10.3389/fnevo.2012.00012

Petrides, M. (2014). *Neuroanatomy of Language Regions of the Human Brain*. Amsterdam: Academic Press. https://doi.org/10.1016/C2011-0-07354-4

Phillips-Silver, J., Aktipis, C. A., & Bryant, G. A. (2010). The Ecology of Entrainment: Foundations of Coordinated Rhythmic Movement. *Music Perception*, *28*(1), 3–14. https://doi.org/10.1525/mp.2010.28.1.3

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, *6*, 26. https://doi.org/10.3389/fnhum.2012.00026

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science*, *308*, 1430–1430. https://doi.org/10.1126/science.1110922

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: auditory encoding of rhythmic movement. *Cognition*, *105*(3), 533–546. https://doi.org/10.1016/j.cognition.2006.11.006

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, *95*(2), 201–236. https://doi.org/10.1016/j.cognition.2004.08.004

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: four cornerstones* (pp. 173–189). Hillsdale: Erlbaum.

Poggio, T. (2012). The Levels of Understanding Framework, Revised. *Perception*, *41*(9), 1017–1023. https://doi.org/10.1068/p7299

Polak, R. (2010). Rhythmic Feel as Meter: Non-Isochronous Beat Subdivision in Jembe Music from Mali. *Music Theory Online*, *16*(4), 1–26.

Polak, R. (2017). The lower limit for meter in dance drumming from West Africa. *Empirical Musicology Review*, *12*(3/4), 205–226.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Pöppel, E. (2004). Lost in time: a historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis*, *64*(3), 295–301. https://doi.org/10.1016/S0301-0511(03)00049-8

Potter, D. D., Fenwick, M., Abecasis, D., & Brochard, R. (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. *Cortex*, *45*(1), 103–109. https://doi.org/10.1016/j.cortex.2008.01.004

Povel, D.-J., & Essens, P. J. (1985). Perception of temporal patterns. *Music Perception: An Interdisciplinary Journal*, *2*(4), 411–440. https://doi.org/10.2307/40285311

Pressing, J. (1983). Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and Western tonality. *Studies in Music*, *17*, 38–61.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*(3–4), 262–275. https://doi.org/10.1080/02643290442000095

Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., … Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology*, *27*(1), 121–131. https://doi.org/10.1037/a0031277

Rauschecker, J. (2013). Brain networks for the encoding of emotions in communication sounds of human and nonhuman primates. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (pp. 1–18). Oxford: Oxford University Press. https://doi.org/10.1093/acprof

Rebuschat, P., Rohrmeier, M., Hawkins, J. A., & Cross, I. (Eds.). (2012). *Language and Music as Cognitive Systems*. Oxford, New York: Oxford University Press.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, *11*(4), 426–428. https://doi.org/10.1038/nn2072

Roberts, I. (2012). Comments and a conjecture inspired by Fabb and Halle. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems*

(pp. 51–66). Oxford: Oxford University Press. https://doi.org/10.1093/acprof

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language*, *141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

Rohrmeier, M. (2007). A generative grammar approach to diatonic harmonic structure, (July), 11–13.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53. https://doi.org/10.1080/17459737.2011.573676

Rohrmeier, M., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *83*(2), 164–175. https://doi.org/10.1016/j.ijpsycho.2011.12.010

Rohrmeier, M., & Pearce, M. (2018). Musical Syntax I: Theoretical Perspectives. In R. Bader (Ed.), *Springer Handbook of Systematic Musicology* (pp. 473–486). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-55004-5_25

Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140097-. https://doi.org/10.1098/rstb.2014.0097

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech Rhythm Facilitates Syntactic Ambiguity Resolution: ERP Evidence. *PLoS ONE*, *8*(2), e56000. https://doi.org/10.1371/journal.pone.0056000

Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., & Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hearing Research*, *308*(9), 71–83. https://doi.org/10.1016/j.heares.2013.08.016

Rouault, M., & Koechlin, E. (2018). Prefrontal function and cognitive control: from action to language. *Current Opinion in Behavioral Sciences*, *21*, 106–111. https://doi.org/10.1016/j.cobeha.2018.03.008

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in Cognitive Sciences*, *19*(2),

86–91. https://doi.org/10.1016/j.tics.2014.12.001

Sambin, S., Teichmann, M., de Diego Balaguer, R., Giavazzi, M., Sportiche, D., Schlenker, P., & Bachoud-Lévi, A.-C. (2012). The role of the striatum in sentence processing: Disentangling syntax from working memory in Huntington's disease. *Neuropsychologia*, *50*(11), 2625–2635. https://doi.org/10.1016/j.neuropsychologia.2012.07.014

Sammler, D. (2009). *The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies*. Universität Leipzig.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., … Schulze-Bonhage, A. (2009). Overlap of Musical and Linguistic Syntax Processing: Intracranial ERP Evidence. *Annals of the New York Academy of Sciences*, *1169*(1), 494–498. https://doi.org/10.1111/j.1749-6632.2009.04792.x

Sammler, D., Koelsch, S., & Friederici, A. D. (2011). Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *47*(6), 659–673. https://doi.org/10.1016/j.cortex.2010.04.007

Samuels, B. D. (2011). *Phonological Architecture*. Oxford, New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199694358.001.0001

Scharff, C., & Adam, I. (2013). Neurogenetics of birdsong. *Current Opinion in Neurobiology*, *23*(1), 29–36. https://doi.org/10.1016/j.conb.2012.10.001

Scharff, C., & Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1574), 2124–2140. https://doi.org/10.1098/rstb.2011.0001

Schiffer, A.-M., & Schubotz, R. I. (2011). Caudate nucleus signals for breaches of expectation in a movement observation paradigm. *Frontiers in Human Neuroscience*, *5*, 38. https://doi.org/10.3389/fnhum.2011.00038

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, *21*(9), 1693–1708. https://doi.org/10.1162/jocn.2008.21153

Schmithorst, V. J. (2005). Separate cortical networks involved in music perception: preliminary functional MRI evidence for modularity of music processing. *NeuroImage*, *25*(2), 444–451. https://doi.org/10.1016/j.neuroimage.2004.12.006

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., … Graybiel, A. M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National*

Academy of Sciences, *111*(39), 14253–14258. https://doi.org/10.1073/pnas.1414542111

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Seger, C. A., Spiering, B. J., Sares, A. G., Quraini, S. I., Alpeter, C., David, J., & Thaut, M. H. (2013). Corticostriatal Contributions to Musical Expectancy Perception. *Journal of Cognitive Neuroscience*, *25*(7), 1062–1077. https://doi.org/10.1162/jocn_a_00371

Seifert, U. (1993). *Systematische Musiktheorie und Kognitionswissenschaft. Zur Grundlegung der kognitiven Musikwissenschaft*. Bonn: Verl. für systemat. Musikwiss.

Seifert, U. (2011). Signification and significance: Music, brain, and culture. Comment on "Towards a neural basis of processing musical semantics" by S. Koelsch. *Physics of Life Reviews*, *8*(2), 122–124. https://doi.org/10.1016/j.plrev.2011.05.007

Seifert, U., Verschure, P. F. M. J., Arbib, M. A., Cohen, A. J., Fogassi, L., Fritz, T., … Rickard, N. (2013). Semantics of Internal and External Worlds. *Language, Music, and the Brain*, 203–229.

Shepherd, G. M. (2004). Introduction to Synaptic Circuits. In G. M. Shepherd (Ed.), *The Synaptic Organization of the Brain* (Vol. 6, pp. 1–38). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159561.003.0001

Shettleworth, S. J. (2010). *Cognition, Evolution, and Behavior* (2nd ed.). Oxford, New York: Oxford University Press.

Shettleworth, S. J. (2012). Modularity, comparative cognition and human uniqueness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2794–2802. https://doi.org/10.1098/rstb.2012.0211

Simonyan, K., Horwitz, B., & Jarvis, E. D. (2012). Dopamine regulation of human speech and bird song: A critical review. *Brain and Language*, *122*(3), 142–150. https://doi.org/10.1016/j.bandl.2011.12.009

Singer, A. (1974). The Metrical Structure of Macedonian Dance. *Ethnomusicology*, *18*(3), 379–404.

Slevc, L. R., & Okada, B. M. (2015). Processing structure in language and music: a case for shared reliance on cognitive control. *Psychonomic Bulletin & Review*, *22*(3), 637–652. https://doi.org/10.3758/s13423-014-0712-4

Slevc, L. R., Reitman, J. G., & Okada, B. M. (2013). Syntax in music and language: The role of cognitive control. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3414–3419). Austin, TX: Cognitive Science Society.

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, *16*(2), 374–381. https://doi.org/10.3758/16.2.374

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians. *NeuroImage*, *17*(3), 1613–1622. https://doi.org/10.1006/nimg.2002.1288

Snyder, B. (2016). *Memory for Music*. (S. Hallam, I. Cross, & M. Thaut, Eds.), *The Oxford Handbook of Music Psychology* (2nd ed.). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.15

Spada, D., Verga, L., Iadanza, A., Tettamanti, M., & Perani, D. (2014). The auditory scene: An fMRI study on melody and accompaniment in professional pianists. *NeuroImage*, *102*(P2), 764–775. https://doi.org/10.1016/j.neuroimage.2014.08.036

Sprouse, J., & Lau, E. F. (2013). Syntax and the brain. In M. den Dikken (Ed.), *The Cambridge Handbook of Generative Syntax* (pp. 971–1005). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511804571.033

Steedman, M. (1996). The Blues and the Abstract Truth: Music and Mental Models. In J. Oakhill & A. Garnham (Eds.), *Mental Models in Cognitive Science: Essays in Honour of Phil Johnson-Laird* (pp. 305–318). Hove: Psychology Press.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*(5), 1169–1178. https://doi.org/10.1093/cercor/bhm149

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2005). Emotional Processing of Harmonic Expectancy Violations. *Annals of the New York Academy of Sciences*, *1060*, 457–461. https://doi.org/10.1196/annals.1360.055

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Stowe, L. A., Paans, A. M. ., Wijers, A. A., & Zwarts, F. (2004). Activations of "motor" and other non-language structures during sentence comprehension. *Brain and Language*, *89*(2), 290–299. https://doi.org/10.1016/S0093-934X(03)00359-6

Sun, L., Liu, F., Zhou, L., & Jiang, C. (2018). Musical training modulates the early but not the late stage of rhythmic syntactic processing. *Psychophysiology*, *55*, e12983.

https://doi.org/10.1111/psyp.12983

Sundberg, J., & Lindblom, B. (1991). Generative Theories for Describing Musical Structure. In P. Howell, R. West, & I. Cross (Eds.), *Representing Musical Structure* (pp. 245–272). London: Academic Press.

Szalisznyó, K., Silverstein, D., Teichmann, M., Duffau, H., & Smits, A. (2017). Cortico-striatal language pathways dynamically adjust for syntactic complexity: A computational study. *Brain and Language*, *164*, 53–62. https://doi.org/10.1016/j.bandl.2016.08.005

Tanimoto, S. L. (1990). *The Elements of Artificial Intelligence Using Common Lisp* (2nd ed.). New York: Computer Science Press.

Teichmann, M., Dupoux, E., Cesaro, P., & Bachoud-Lévi, A.-C. (2008). The role of the striatum in sentence processing: Evidence from a priming study in early stages of Huntington's disease. *Neuropsychologia*, *46*(1), 174–185. https://doi.org/10.1016/j.neuropsychologia.2007.07.022

Teichmann, M., Dupoux, E., Kouider, S., & Bachoud-Lévi, A.-C. (2006). The Role of the Striatum in Processing Language Rules: Evidence from Word Perception in Huntington's Disease. *Journal of Cognitive Neuroscience*, *18*(9), 1555–1569. https://doi.org/10.1162/jocn.2006.18.9.1555

Teichmann, M., Dupoux, E., Kouider, S., Brugières, P., Boissé, M.-F., Baudic, S., … Bachoud-Lévi, A.-C. (2005). The role of the striatum in rule application: the model of Huntington's disease at early stage. *Brain : A Journal of Neurology*, *128*, 1155–1167. https://doi.org/10.1093/brain/awh472

Teichmann, M., Rosso, C., Martini, J.-B., Bloch, I., Brugières, P., Duffau, H., … Bachoud-Lévi, A.-C. (2015). A cortical-subcortical syntax pathway linking Broca's area and the striatum. *Human Brain Mapping*, *36*(6), 2270–2283. https://doi.org/10.1002/hbm.22769

Teki, S., Grube, M., & Griffiths, T. D. (2012). A Unified Model of Time Perception Accounts for Duration-Based and Beat-Based Timing Mechanisms. *Frontiers in Integrative Neuroscience*, *5*, 90. https://doi.org/10.3389/fnint.2011.00090

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing. *Journal of Neuroscience*, *31*(10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011

Temperley, D. (2000). Meter and Grouping in African Music: A View from Music Theory. *Ethnomusicology*, *44*(1), 65–96.

ten Cate, C. (2017). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin & Review*, *24*(1), 91–96.

https://doi.org/10.3758/s13423-016-1091-9

ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1984–1994. https://doi.org/10.1098/rstb.2012.0055

Terzi, A., Papapetropoulos, S., & Kouvelas, E. D. (2005). Past tense formation and comprehension of passive sentences in Parkinson's disease: Evidence from Greek. *Brain and Language*, *94*(3), 297–303. https://doi.org/10.1016/j.bandl.2005.01.005

Thaut, M. H., Demartin, M., & Sanes, J. N. (2008). Brain networks for integrative rhythm formation. *PLoS ONE*, *3*(5), 1–10. https://doi.org/10.1371/journal.pone.0002312

Thaut, M. H., Miltner, R., Lange, H. W., Hurt, C. P., & Hoemberg, V. (1999). Velocity modulation and rhythmic synchronization of gait in Huntington's disease. *Movement Disorders*, *14*(5), 808–819. https://doi.org/10.1002/1531-8257(199909)14:5<808::AID-MDS1014>3.0.CO;2-J

Thompson-Schill, S., Hagoort, P., Dominey, P. F., Honing, H., Koelsch, S., Ladd, D. R., … Steedman, M. (2013). Multiple Levels of Structure in Language and Music. In M. A. Arbib (Ed.), *Language, music* (pp. 289–303). Cambridge, MA: The MIT Press.

Thompson, W. F., & Balkwill, L.-L. (2010). Cross-Cultural Similarities and Differences. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 755–788). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0027

Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. *Topics in Cognitive Science*, *4*(4), 568–584. https://doi.org/10.1111/j.1756-8765.2012.01209.x

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the Inferior Frontal Cortex in Musical Priming. *Annals of the New York Academy of Sciences*, *999*, 209–211. https://doi.org/10.1196/annals.1284.031

Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., & von Cramon, D. Y. (2006). Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, *31*(4), 1771–1782. https://doi.org/10.1016/j.neuroimage.2006.02.028

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift Für Tierpsychologie*, *20*, 410–433.

Tolleson, C. M., Dobolyi, D. G., Roman, O. C., Kanoff, K., Barton, S., Wylie, S. A., …

Claassen, D. O. (2015). Dysrhythmia of timed movements in Parkinson′s disease and freezing of gait. *Brain Research*, *1624*(5), 222–231. https://doi.org/10.1016/j.brainres.2015.07.041

Tomasello, M. (1996). Do Apes Ape? In C. M. Heyes (Ed.), *Social Learning in Animals: The Roots of Culture* (pp. 319–346). London: Academic Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Toussaint, G. (2003). Classification and Phylogenetic Analysis of African Ternary Rhythm Timelines. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science* (pp. 25–36). Granada.

Toussaint, G. (2005). The Euclidean Algorithm Generates Traditional Musical Rhythms. In *Bridges: Mathematical Connections in Art, Music and Science* (pp. 47–56). Banff.

Trost, W., Labbé, C., & Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia*, *96*, 96–110. https://doi.org/10.1016/j.neuropsychologia.2017.01.004

Trost, W., & Vuilleumier, P. (2013). Rhythmic entrainment as a mechanism for emotion induction by music. In T. Cochrane, B. Fantini, & K. R. Scherer (Eds.), *The Emotional Power of Music* (pp. 213–225). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654888.003.0016

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-Analysis of the Functional Neuroanatomy of Single-Word Reading: Method and Validation. *NeuroImage*, *16*(3), 765–780. https://doi.org/10.1006/nimg.2002.1131

Tyler, L. K., & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1037–1054. https://doi.org/10.1098/rstb.2007.2158

Uddén, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1598), 2023–2032. https://doi.org/10.1098/rstb.2012.0009

Ullman, M. T. (2006). Is Broca's Area Part of a Basal Ganglia Thalamocortical Circuit? *Cortex*, *42*(4), 480–485. https://doi.org/10.1016/S0010-9452(08)70382-4

Ullman, M. T., & Pierpont, E. I. (2005). Specific Language Impairment is not Specific to Language: the Procedural Deficit Hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition*, *146*, 172–184. https://doi.org/10.1016/j.cognition.2015.09.013

Vazan, P., & Schober, M. F. (2004). Detecting and Resolving Metrical Ambiguity in a Rock Song Upon Multiple Rehearings. In *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 426–432). Adelaide: Causal Productions.

Vernes, S. C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin & Review*, *24*(1), 111–117. https://doi.org/10.3758/s13423-016-1060-3

Vogeley, K., & Roepstorff, A. (2009). Contextualising culture and social cognition. *Trends in Cognitive Sciences*, *13*(12), 511–516. https://doi.org/10.1016/j.tics.2009.09.006

Vuust, P., Gebauer, L. K., & Witek, M. A. G. (2014). Neural Underpinnings of Music: The Polyrhythmic Brain. In H. Merchant & V. de Lafuente (Eds.), *Neurobiology of interval timing* (pp. 339–356). New York: Springer. https://doi.org/10.1007/978-1-4939-1782-2_18

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing... Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *NeuroImage*, *31*(2), 832–841. https://doi.org/10.1016/j.neuroimage.2005.12.037

Vuust, P., Wallentin, M., Mouridsen, K., Østergaard, L., & Roepstorff, A. (2011). Tapping polyrhythms in music activates language areas. *Neuroscience Letters*, *494*(3), 211–216. https://doi.org/10.1016/j.neulet.2011.03.015

Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*, 1111. https://doi.org/10.3389/fpsyg.2014.01111

Wallin, N. L. (1991). *Biomusicology. Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music*. Stuyvesant, NY: Pendragon Press.

Ward, J. (2012). *The Student's Guide to Social Neuroscience* (1st ed.). Hove: Psychology Press.

Watkins, K. E., & Jenkinson, N. (2016). The Anatomy of the Basal Ganglia. In *Neurobiology of Language* (pp. 85–94). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00008-0

Wiener, M., Matell, M. S., & Coslett, H. B. (2011). Multiple Mechanisms for Temporal Processing. *Frontiers in Integrative Neuroscience*, *5*, 31. https://doi.org/10.3389/fnint.2011.00031

Wiener, M., Turkeltaub, P., & Coslett, H. B. (2010). The image of time: A voxel-wise meta-analysis. *NeuroImage*, *49*(2), 1728–1740.

https://doi.org/10.1016/j.neuroimage.2009.09.064

Wilkins, R. W. (2018). Network Neuroscience: An Introduction to Graph Theory Network-Based Techniques for Music and Brain Imaging Research. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford Handbook of Music and the Brain* (pp. 1–25). Oxford: Oxford University Press. https://doi.org/10.110.1093/oxfordhb/9780198804123.013.5

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved Sequence Processing in Primate Frontal Cortex. *Trends in Neurosciences*, *40*(2), 72–82. https://doi.org/10.1016/j.tins.2016.11.004

Winkler, I. (2007). Interpreting the Mismatch Negativity. *Federation of European Psychophysiology Societies*, *21*(3–4), 147–163. https://doi.org/10.1027/0269-8803.21.3.147

Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, Body-Movement and Pleasure in Groove Music. *PLoS ONE*, *9*(4), e94446. https://doi.org/10.1371/journal.pone.0094446

Witek, M. A. G., Kringelbach, M. L., & Vuust, P. (2015). Musical rhythm and affect. Comment on "The quartet theory of human emotions: An integrative and neurofunctional model" by S. Koelsch et al. *Physics of Life Reviews*, *13*, 92–94. https://doi.org/10.1016/j.plrev.2015.04.029

Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research*, *1130*, 167–180. https://doi.org/10.1016/j.brainres.2006.10.052

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Zaccarella, E., Schell, M., & Friederici, A. D. (2017). Reviewing the functional basis of the syntactic Merge mechanism for language: A coordinate-based activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews*, *80*, 646–656. https://doi.org/10.1016/j.neubiorev.2017.06.011

Zatorre, R. J., & Peretz, I. (2001). The biological foundations of music. *Annals of the New York Academy of Sciences*, *930*(1), ix–x. https://doi.org/10.1111/j.1749-6632.2001.tb05720.x

Zentall, T. R. (1996). An Analysis of Imitative Learning in Animals. In C. M. Heyes (Ed.), *Social Learning in Animals: The Roots of Culture* (pp. 221–243). London: Academic Press. https://doi.org/10.1016/B978-012273965-1/50012-1

List of abbreviations

This list includes abbreviations used in cognitive neuroscience research, evolutionary research, and theoretical research. The theoretical concepts, hypotheses, and theories are written in italics.

a-	anterior
ACC	Anterior cingulate cortex
[ACC]	*Accusative*
ACT	*Activation*
ACT-R	*Adaptive Control of Thought-Rational*
AF	Arcuate fasciculus
AGL	*Artificial grammar learning*
AgrOP	*Object agreement phrase*
AgrSP	*Subject agreement phrase*
ALE	Activation likelihood estimation
ASAP	*Action simulation for auditory prediction*
BA	Brodmann area
BAT	*Beat alignment tests*
BG	Basal ganglia
BODB	*Brain Operation Database*
BOP	*Brain operation principle*
BPM	*Beats per minute*
BPS	*Beat perception and synchronization*
CAUD	Caudate
CBGT	Cortico-basal ganglia-thalamocortical
cFWE	Cluster-level family-wise error
cl-	caudolateral
COG	*Cognitive process*
CR	*Computational-representational*
d-	dorsal
d	*dominant*
D	Dorsal stream / pathway
D	Dopamine receptor

DA	Dopaminergic
Det	*Determiner*
DIVA	*Directions Into Velocities of Articulators*
dl-	dorsolateral
DLPFC	Dorsolateral prefrontal cortex
DLT	*Dependency Locality Theory*
dm-	dorsomedial
DP	*Determiner phrase*
DR	*Dominant region*
DTI	Diffusion tensor imaging
EEG	Electroencephalography
EFCS	Extreme fiber capsule system
ELAN	Early left anterior negativity
ERAN	Early right anterior negativity
ERP	Event-related potential
FC	Frontal cortex
fMRI	Functional magnetic resonance imaging
FOP	Frontal operculum
G	*Grammar*
GAE	*Gradual Audiomotor Evolution*
Glu	Glutamergic
GMP	*Granularity Mismatch Problem*
GODIVA	*Gradient Order DIVA*
GPe	Globus pallidus external
GPi	Globus pallidus internal
GPR	*Grouping preference rule*
GSTH	*Generative Syntax of Tonal Harmony*
GTTM	*A Generative Theory of Tonal Music*
GWFR	*Grouping well-formedness rule*
HC	Hippocampal cortex
HD	Huntington's disease
IBI	*Inter-beat interval*
IFC	Inferior frontal cortex
IFG	Inferior frontal gyrus

IFOF	Longitudinal inferior-fronto-occipital fasciculus
IOI	*Inter-onset interval*
ITI	*Inter-tap interval*
LAN	Left anterior negativity
LCA	Last common ancestor
[M]	*Masculine*
MC	Motor cortex
MD	Medialis dorsalis
MDpc	Medialis dorsalis pars pavocellularis
MEG	Magnetoencephalography
MLF	Middle longitudinal fasciculus
MMN	Mismatch Negativity
MNI	Montreal Neurological Institute
MPR	*Metrical preference rule*
MSN	Medium spiny neuron
MTG	Middle temporal gyrus
MUC	*Memory unification control*
MWFR	*Metrical well-formedness rule*
N	*Noun*
[N]	*Neuter*
NAcc	Nucleus accumbens
NB	*Neurobiological*
[NOM]	*Nominative*
NP	*Noun phrase*
OC	Occipital cortex
OFC	Orbitofrontal cortex
OIP	*Ontological Incommensurability Problem*
p-	posterior
-p	*parallel*
P	*Phrase*
PC	Parietal cortex
PD	Parkinson's disease
PET	Positron-emission tomography
pm-	posteromedial

PM	*Perceptual model*
PMC	Premotor cortex
PP	*Prepositional phrase*
PPC	Posterior parietal cortex
PUT	Putamen
RC	*Relative clause*
rd-	rostrodorsal
rl-	rostrolateral
ROI	Region of interest
s	*subdominant*
S	*Sentence*
S	Striosome
SC	Somatosensory cortex
SED	*Summaries of Empirical Data*
SLF	Superior longitudinal fasciculus
SLI	Specific language impairment
SMA	Supplementary motor area
SNc	Substantia nigra pars compacta
SNr	Substantia nigra pars reticulate
SPM	Statistical parametric mapping
SR	*Subdominant region*
SSIRH	*Shared syntactic integration resource hypothesis*
SSR	*Summaries of Simulation Results*
STG	Superior temporal gyrus
STN	Subthalamic nucleus
STS	Superior temporal sulcus
t	*tonic*
TC	Temporal cortex
TPS	*Tonal pitch space*
TR	*Tonic region*
UF	Uncinate fasciculus
v-	ventral
V	*Verb*
V	Ventral stream / pathway

VApc	Ventralis anterior pars parvocellularis
vl-	ventrolateral
VLm	Ventralis lateralis pars medialis
VLo	Ventralis lateralis pars oralis
VLPFC	Ventrolateral prefrontal cortex
VLSM	Voxel-based lesion-symptom mapping
VMPFC	Ventromedial prefrontal cortex
VP	*Verbal phrase*
VP	Ventral pallidum
VS	Ventral striatum
WCST	*Wisconsin Card Sorting Task*
WFC	*Well-formedness constraints*
ZEBrA	*Zebra Finch Expression Brain Atlas*

Appendix A: Output files of the ALE meta-analyses

Appendix A contains information provided by GingerALE output files after conducting the ALE meta-analyses presented in the current thesis. This includes information concerning analysis parameter settings and more detailed cluster descriptions. This information is provided to ease other researchers to replicate the current meta-analyses.

1 An ALE meta-analysis of tonal-harmonic syntactic processing

Mask:

Dimensions	= 77x96x79
Number of within-brain voxels	= 264227
Mask Size	= Less Conservative (Larger)

Foci:

Coordinate System	= MNI
File of foci coordinates	= Syntax_melody-harmony_MNI_2018_08_16.txt
Number of foci	= 193
Number of experiments	= 17
Total number of subjects	= 324

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):

File of ALE voxels	= Syntax_melody-harmony_MNI_2018_08_16_ALE.nii
FWHM minimum value	= 8.706421297150076
FWHM median value	= 9.437333897527274
FWHM maximum value	= 10.002568195473065
Minimum ALE score	= 7.675972E-35
Maximum ALE score	= 0.02577522

P Values: Eickhoff (HBM, 2009)

File of P values	= Syntax_melody-harmony_MNI_2018_08_16_P.nii
Minimum P value	= 1.1187414E-9

Thresholding:

Threshold Method	= Cluster-level Inference

Thresholding Value = 0.05
Thresholding Permutations = 1000
Cluster-Forming Method = Uncorrected P value
Cluster-Forming Value = 0.001
Volume > Threshold = 11296 mm^3
Chosen min. cluster size = 832 mm^3
Thresholded ALE image = Syntax_melody-harmony_MNI_2018_08_16_ALE_C05_1k.nii

Cluster Analysis:
#1: 7528 mm^3 from (28,8,-16) to (58,42,28) centered at (45.2,22.5,1.9)
Extrema: (Labels are within +/-5mm)
0.02577522 at (34,24,0) Right Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 13
0.022346536 at (52,18,8) Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 44
0.020763263 at (48,16,-2) Right Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 13
0.018753408 at (44,34,2) Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 13
0.014828126 at (44,20,16) Right Cerebrum.Frontal Lobe.Middle Frontal Gyrus.Gray Matter.Brodmann area 46
0.013811457 at (38,12,-2) Right Cerebrum.Sub-lobar.Claustrum.Gray Matter.*
0.013333128 at (50,22,-12) Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 47
0.011143876 at (48,22,26) Right Cerebrum.Frontal Lobe.Middle Frontal Gyrus.Gray Matter.Brodmann area 9
Labels: (Nearest gray within +/-5mm)
1792mm Right Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 13
976mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 47
880mm Right Cerebrum.Sub-lobar.Claustrum.Gray Matter.*
744mm Right Cerebrum.Frontal Lobe.Precentral Gyrus.Gray Matter.Brodmann area 44
648mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 46
616mm Right Cerebrum.Frontal Lobe.Middle Frontal Gyrus.Gray Matter.Brodmann area 46
448mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 45
440mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 13
400mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 44

184mm Right Cerebrum.Sub-lobar.Insula.Gray Matter.*
136mm Right Cerebrum.Frontal Lobe.Middle Frontal Gyrus.Gray Matter.Brodmann area 9
104mm No Gray Matter found
48mm Right Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 47
40mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.*
40mm Right Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 9
24mm Right Cerebrum.Sub-lobar.Inferior Frontal Gyrus.Gray Matter.Brodmann area 45
8mm Right Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 45

#2: 2696 mm^3 from (-60,2,-16) to (-42,18,6) centered at (-50.4,9.9,-5.6)
Extrema: (Labels are within +/-5mm)
0.017942585 at (-52,6,-10) Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
0.016120544 at (-52,10,0) Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
Labels: (Nearest gray within +/-5mm)
912mm Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
840mm Left Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 13
352mm Left Cerebrum.Frontal Lobe.Precentral Gyrus.Gray Matter.Brodmann area 44
312mm Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 38
192mm Left Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.Brodmann area 47
72mm Left Cerebrum.Sub-lobar.Insula.Gray Matter.*
16mm Left Cerebrum.Frontal Lobe.Inferior Frontal Gyrus.Gray Matter.*

#3: 1072 mm^3 from (48,-42,-6) to (64,-30,12) centered at (57.7,-33.7,4.4)
Extrema: (Labels are within +/-5mm)
0.018720051 at (60,-34,8) Right Cerebrum.Temporal Lobe.Middle Temporal Gyrus.Gray Matter.Brodmann area 22
0.012114726 at (52,-32,0) Right Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
0.011361487 at (58,-40,-4) Right Cerebrum.Temporal Lobe.Middle Temporal Gyrus.Gray Matter.Brodmann area 22

Labels: (Nearest gray within +/-5mm)

312mm Right Cerebrum.Temporal Lobe.Middle Temporal Gyrus.Gray Matter.Brodmann area 22

296mm Right Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22

176mm Right Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 41

120mm Right Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 42

96mm Right Cerebrum.Temporal Lobe.Middle Temporal Gyrus.Gray Matter.*

56mm Right Cerebrum.Temporal Lobe.Middle Temporal Gyrus.Gray Matter.Brodmann area 21

8mm No Gray Matter found

8mm Right Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 22

Experiment Table:

[3 4 0 4 0 2 2 3 0 1 2 1 0 1 1 3 2]

[0 0 0 2 0 1 2 2 0 1 0 0 0 1 0 1 1]

[2 1 0 2 0 0 1 1 0 0 0 0 0 0 0 0 0]

Contributors to cluster #1

3 foci from Bianco, 2016: incongruent > congruent

4 foci from Cheung, 2018: ungrammatical > grammatical

4 foci from Fedorenko, 2011: intact > scrambled

2 foci from Fujisawa, 2011: harmony > white noise

2 foci from Koelsch S, 2002: modulation > in-key

3 foci from Koelsch S, 2005: Irregular > Regular

1 foci from Levitin, 2003: normal > scrambled

2 foci from Musso, 2015: structural deviant > well-formed

1 foci from Oechslin, 2013: main effect of transregression

1 foci from Seger, 2013: cadences > baseline

1 foci from Spada, 2014: altered melody > correct melody

3 foci from Tillmann, 2003: UC > RC

2 foci from Tillmann, 2005: less related > related

Contributors to cluster #2

2 foci from Fedorenko, 2011: intact > scrambled

1 foci from Fujisawa, 2011: harmony > white noise

2 foci from Koelsch S, 2002: modulation > in-key

2 foci from Koelsch S, 2005: Irregular > Regular

1 foci from Levitin, 2003: normal > scrambled

1 foci from Seger, 2013: cadences > baseline

1 foci from Tillmann, 2003: UC > RC

1 foci from Tillmann, 2005: less related > related

Contributors to cluster #3

2 foci from Bianco, 2016: incongruent > congruent

1 foci from Cheung, 2018: ungrammatical > grammatical

2 foci from Fedorenko, 2011: intact > scrambled

1 foci from Koelsch S, 2002: modulation > in-key

1 foci from Koelsch S, 2005: Irregular > Regular

2 An ALE meta-analysis of rhythmic syntactic processing

Mask:

Dimensions	= 77x96x79
Number of within-brain voxels	= 264227
Mask Size	= Less Conservative (Larger)

Foci:

Coordinate System	= MNI
File of foci coordinates	= Meta-analysis_Rhythm_2018-12-30.txt
Number of foci	= 213
Number of experiments	= 18
Total number of subjects	= 281

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):

File of ALE voxels	= Meta-analysis_Rhythm_2018-12-30_ALE.nii
FWHM minimum value	= 8.881953893794925
FWHM median value	= 9.755397007122182

FWHM maximum value = 10.362276299774612
Minimum ALE score = 2.5330834E-33
Maximum ALE score = 0.02207786

P Values: Eickhoff (HBM, 2009)
File of P values = Meta-analysis_Rhythm_2018-12-30_P.nii
Minimum P value = 4.1158213E-8

Thresholding:
Threshold Method = Cluster-level Inference
Thresholding Value = 0.05
Thresholding Permutations = 1000
Cluster-Forming Method = Uncorrected P value
Cluster-Forming Value = 0.001
Volume > Threshold = 9624 mm^3
Chosen min. cluster size = 696 mm^3
Thresholded ALE image = Meta-analysis_Rhythm_2018-12-30_ALE_C05_1k.nii

Cluster Analysis:
#1: 4144 mm^3 from (-10,-8,48) to (10,20,72) centered at (.4,3,59.1)
Extrema: (Labels are within +/-5mm)
0.019667547 at (-2,-4,58) Left Cerebrum.Frontal Lobe.Medial Frontal Gyrus.Gray Matter.Brodmann area 6
0.015092112 at (2,16,54) Left Cerebrum.Frontal Lobe.Superior Frontal Gyrus.Gray Matter.Brodmann area 6
0.012789007 at (2,-4,70) Left Cerebrum.Frontal Lobe.Medial Frontal Gyrus.Gray Matter.Brodmann area 6
Labels: (Nearest gray within +/-5mm)
2200mm Left Cerebrum.Frontal Lobe.Medial Frontal Gyrus.Gray Matter.Brodmann area 6
864mm Right Cerebrum.Frontal Lobe.Medial Frontal Gyrus.Gray Matter.Brodmann area 6
640mm Left Cerebrum.Frontal Lobe.Superior Frontal Gyrus.Gray Matter.Brodmann area 6
424mm Right Cerebrum.Frontal Lobe.Superior Frontal Gyrus.Gray Matter.Brodmann area 6
16mm Left Cerebrum.Frontal Lobe.Medial Frontal Gyrus.Gray Matter.Brodmann area 32

#2: 2216 mm^3 from (-38,-72,-42) to (-24,-60,-18) centered at (-30.9,-65.3,-28.8)
Extrema: (Labels are within +/-5mm)
0.02207786 at (-30,-64,-26) Left Cerebellum.Anterior Lobe.Culmen.Gray Matter.*
Labels: (Nearest gray within +/-5mm)
648mm Left Cerebellum.Anterior Lobe.Culmen.Gray Matter.*
616mm Left Cerebellum.Posterior Lobe.Cerebellar Tonsil.Gray Matter.*
312mm Left Cerebellum.Posterior Lobe.Declive.Gray Matter.*
232mm Left Cerebellum.Posterior Lobe.Pyramis.Gray Matter.*
144mm Left Cerebellum.Posterior Lobe.Tuber.Gray Matter.*
144mm Left Cerebellum.Posterior Lobe.Uvula.Gray Matter.*
104mm Left Cerebellum.Anterior Lobe.*.Gray Matter.*
16mm Left Cerebellum.Posterior Lobe.Inferior Semi-Lunar Lobule.Gray Matter.*

#3: 1896 mm^3 from (-58,-12,-10) to (-46,14,6) centered at (-52.7,3.5,-1)
Extrema: (Labels are within +/-5mm)
0.017584482 at (-54,8,0) Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
0.015968941 at (-52,-8,-2) Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
Labels: (Nearest gray within +/-5mm)
1176mm Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.Brodmann area 22
384mm Left Cerebrum.Frontal Lobe.Precentral Gyrus.Gray Matter.Brodmann area 44
168mm Left Cerebrum.Sub-lobar.Insula.Gray Matter.Brodmann area 13
112mm Left Cerebrum.Temporal Lobe.Superior Temporal Gyrus.Gray Matter.*
56mm Left Cerebrum.Sub-lobar.Insula.Gray Matter.*

#4: 1368 mm^3 from (14,4,-2) to (28,14,10) centered at (22.6,9,4)
Extrema: (Labels are within +/-5mm)
0.019573327 at (24,10,6) Right Cerebrum.Sub-lobar.Lentiform Nucleus.Gray Matter.Putamen
Labels: (Nearest gray within +/-5mm)
1152mm Right Cerebrum.Sub-lobar.Lentiform Nucleus.Gray Matter.Putamen
184mm Right Cerebrum.Sub-lobar.Caudate.Gray Matter.Caudate Body
32mm Right Cerebrum.Sub-lobar.Lentiform Nucleus.Gray Matter.Lateral Globus Pallidus

Experiment Table:

[0 5 1 0 0 0 0 2 1 1 1 1 1 1 1 0 1 2]

[0 0 0 1 0 0 0 1 1 1 1 1 1 1 0 0 0 0]

[0 1 1 2 0 2 0 0 0 1 0 2 0 0 0 0 0 0]

[0 0 1 0 1 1 0 0 1 0 0 1 0 0 1 0 0 0]

Contributors to cluster #1

5 foci from Thaut, 2008: Polyrhythmic tapping > listening

1 foci from Kung, 2013: beat finding > listen isochronous

2 foci from Chen, 2008b: Covariation with complexity

1 foci from Chen, 2008a: Anticipatory listening > Silence (Group 1)

1 foci from Chen, 2008a: Anticipatory listening > Silence (Group 2)

1 foci from Oullier, 2005: ImSyncopation > ImSynchronization

1 foci from Mayville, 2002: Syncopation > Synchronization

1 foci from Jantzen, 2007: Syncopation > Synchronization

1 foci from Jantzen, 2005: Syncopation > Synchronization

1 foci from Jantzen, 2002: Syncopation > rest (pre-practice)

1 foci from Chapin, 2010: Auditory attend 2 > rest

2 foci from Bengtsson, 2009: Rhythmic > Random

Contributors to cluster #2

1 foci from Konoike, 2012: Rhythm Encoding auditory > Number Encoding auditory

1 foci from Chen, 2008b: Covariation with complexity

1 foci from Chen, 2008a: Anticipatory listening > Silence (Group 1)

1 foci from Chen, 2008a: Anticipatory listening > Silence (Group 2)

1 foci from Oullier, 2005: ImSyncopation > ImSynchronization

1 foci from Mayville, 2002: Syncopation > Synchronization

1 foci from Jantzen, 2007: Syncopation > Synchronization

1 foci from Jantzen, 2005: Syncopation > Synchronization

Contributors to cluster #3

1 foci from Thaut, 2008: Polyrhythmic tapping > listening

1 foci from Kung, 2013: beat finding > listen isochronous

2 foci from Konoike, 2012: Rhythm Encoding auditory > Number Encoding auditory

2 foci from Grahn, 2007: Metrically simple > Metrically complex & non-metric

1 foci from Chen, 2008a: Anticipatory listening > Silence (Group 2)

2 foci from Mayville, 2002: Syncopation > Synchronization

Contributors to cluster #4

1 foci from Kung, 2013: beat finding > listen isochronous

1 foci from Grahn, 2009: Beat > Nonbeat

1 foci from Grahn, 2007: Metrically simple > Metrically complex & non-metric

1 foci from Chen, 2008a: Anticipatory listening > Silence (Group 1)

1 foci from Mayville, 2002: Syncopation > Synchronization

1 foci from Jantzen, 2002: Syncopation > rest (pre-practice)

Appendix B: Illustration of Agbekor dance

In PART II (Section 12.2, pp. 102), the structure of African drum ensemble music was described to develop a computational-representational theory of rhythmic syntax which is also applicable to non-Western music tradition. This appendix illustrates main aspects of this analysis of an Agbekor performance. It focuses on the illustration of dance figures which played an important role to analyze metrical and grouping structure in PART II (Section 12.2, pp. 102). The video material was recorded by Jason Aryeh. It is available from YouTube (https://www.youtube.com/watch?v=7NUjDKOiyas) and was published on September 7th, 2013 and retrieved on October 28th, 2018. The analysis was carried out with ELAN software which is an open-source annotation software (https://tla.mpi.nl/tools/tla-tools/elan/) developed and made available by Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. The figures are edited by GIMP software which enables picture editing and Microsoft PowerPoint.

Figure 2.1 Agbekor Ensemble (Ewe, Ghana). This ensemble consists of five instrumentalists and four dancers. There are five instruments: Axatse (top left), Gankogui (top right), and three drums. The picture was edited so that the identification of the performers is not possible.

An Agbekor performance consists of singing, drumming[70], and dancing. The timeline pattern is played on the Gankogui. The performance which I analyzed has a total length of approximately nine minutes (from 00:08.000 to 09:12.000)[71] and is made up of two parts, i.e., a slow part (00:08.000 - 04:33.360) and a fast part (04:33.360 - 09:12.000), four sections as well as different sub-sections: an opening section (00:08.000 - 01:29.000), a main slow section (01:29.000 - 04:33.360), a main fast section (04:33.360 - 07:59.420), and a closing section (07:59.420 - 09:12.000). The opening section begins with singing and drumming, then the dancers enter the 'stage' to perform the two main sections. In the closing section, the dancers bow down and leave the 'stage'. The performance ends with singing and drumming. The drumming stretches out over the entire performance and the singing accompanies the entire performance from time to time. The discussions in the current thesis referred to the first thirty seconds of the main slow section (01:29.00 - 01:58.080). The analysis window is shown in Figure 2.2.

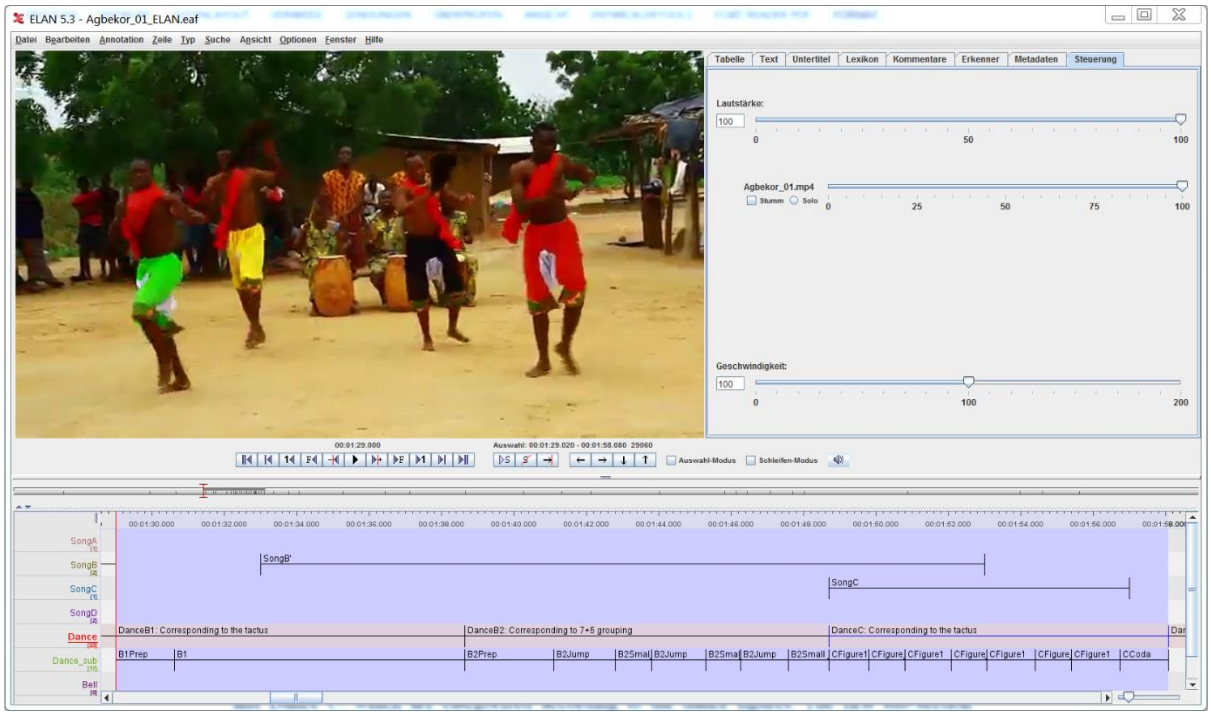


Figure 2.2 The main slow section in the analysis window 01:29.00 - 01:58.080 as displayed in ELAN.

The main slow section consists of three sub-sections, i.e., i.e., Dance B1, Dance B2, and Dance C, which are categorized according to the dance figures. The first sub-section

[70] Drumming also includes playing the rattle (Axatse) and the bell (Gankogui).
[71] "00:08.000" means 0 minute 8 seconds 000 millisecond.

(Dance B1 in Figure 2.2; 01:29.000 - 01:38.640) includes steps and arm movements which strongly corresponds to the primary beats. Especially, through the steps from left to right and vice versa, the primary beats are clearly identifiable (Figure 2.3 Dance figure B1). The dance figure in the second sub-section (B2 in Figure 2.2; 01:38.640 - 01:48.704) does not strongly corresponds to the primary beats. Rather, the movements make the 7+5 grouping clear. The first large dance movement corresponds to seven beats at the lowest subdivision level and the small knee movements correspond to five beats (Figure 2.3 Dance figure B2). The third subsection (C in Figure 2.2; 01:48.704 - 01:58.080) again includes dance steps strongly corresponding to the primary beats. In this way, from sub-section to sub-section, dance movements put emphasis on different aspects of the piece's rhythmic structure, namely metrical and grouping structure. Because the illustration of this appendix is limited, it is recommended to watch the video to get closer impression of Agbekor performance.

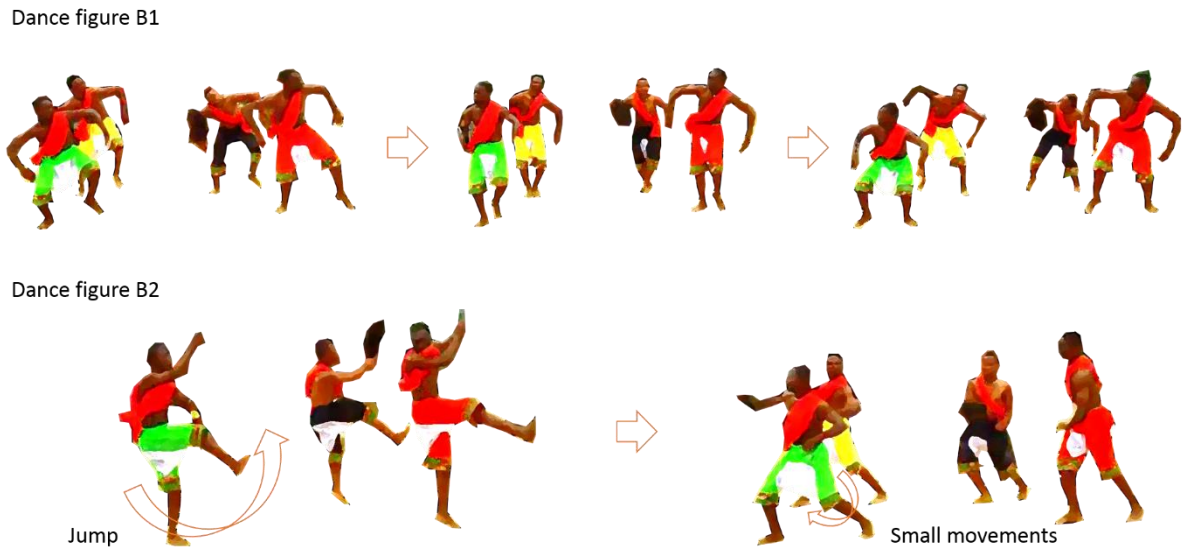


Figure 2.3 Dance figures of the first and second subsection of the main slow section.

Reference

ELAN (Version 5.2) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/